Transition States in the Evolution of the Mammalian Sex Chromosomes

by
Karin Jegalian

B.S. Biochemistry
University of California, Los Angeles (1992)

Submitted to the Department of Biology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
at the
Massachusetts Institute of Technology

September 1997
[February 1997]

Signature of author_____
Department of Biology
September 1997

Certified by_____
David C. Page
Professor, Department of Biology
Thesis Supervisor

Accepted by_____
Frank Solomon
Professor, Department of Biology
Chairman, Committee on Graduate Studies

Acknowledgment

My father's faith in scholarship
and specifically in molecular biology incited this work.

Table of Contents

Transition States in the Evolution of the Mammalian Sex Chromosomes

by

Karin Jegalian

Submitted to the Department of Biology in September 1997
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

ABSTRACT

Over a dozen genes are known in the human which escape X inactivation; most have conserved homologs on the Y chromosome. In almost all cases, the homologous mouse genes are subject to X inactivation and lack Y homologs. Which dosage strategy—expressing two copies of sex chromosome genes per cell, as in human, or expressing one copy per cell, as in mouse—represents a more primitive and which a more derived state?

CpG island methylation was used as a surrogate assay for X inactivation in a diverse array of eutherian mammals. This technique is justified by the example of *ALD*. Methylation of 5' CpG islands can be applied to study allelic inactivation generally.

Case studies of *ZFX/ZFY*, *RPS4X/RPS4Y*, and *SMCX/SMCY* reveal incremental changes toward Y gene degeneration and X inactivation that illustrate broad principles in the evolution of sex chromosomes and show that the broad tendencies can operate idiosyncratically at the level of individual genes. Transition states, instances of incompletion, still exist in the evolution of mammalian sex chromosomes toward wider dosage compensation of the X chromosome and toward decay of the Y chromosome.

Thesis Advisor: David C. Page
Title: Professor of Biology

Chapter 1. Introduction

Dosage of Sex Chromosome Genes

Sex chromosome systems have evolved independently in a broad range of organisms. Though genetically unrelated, analogous sex chromosome systems, shaped by the same fundamental forces, have convergent properties. In the end, a pair of sex chromosomes arise that look different from each other and do not recombine with each other along most of their lengths, where one chromosome is small, heterochromatic, and bereft of genes and the other, the more genetically and cytogenetically normal chromosome, is subject to some system of dosage compensation. The sex chromosomes within a species, however distinctive now, appear to share a common origin, having arisen from a pair of identical chromosomes (autosomes).

The process of sex chromosome differentiation may begin when an allele arises that affects sex determination. When the locus is heterozygous, the organism develops as one sex, the heterogametic sex, male in the case of mammals; when the locus is homozygous, the organism becomes the homogametic sex, female in the case of mammals. Over time, recombination is suppressed around the sex-determining locus. The heterogametic chromosome—called the Y chromosome where the heterogametic sex is male—degenerates; and for the other chromosome—called the X where the homogametic sex is female—a system of dosage compensation develops that equalizes gene expression levels for that chromosome between the two sexes.

Figures 1.1 and 1.2 show, respectively, the human and mouse sex chromosomes. In both species the Y chromosome is small and heterochromatic. The X and Y chromosomes, though morphologically so distinctive, still pair during male meiosis. Recombination occurs in small pseudoautosomal regions (PARs) at the tips

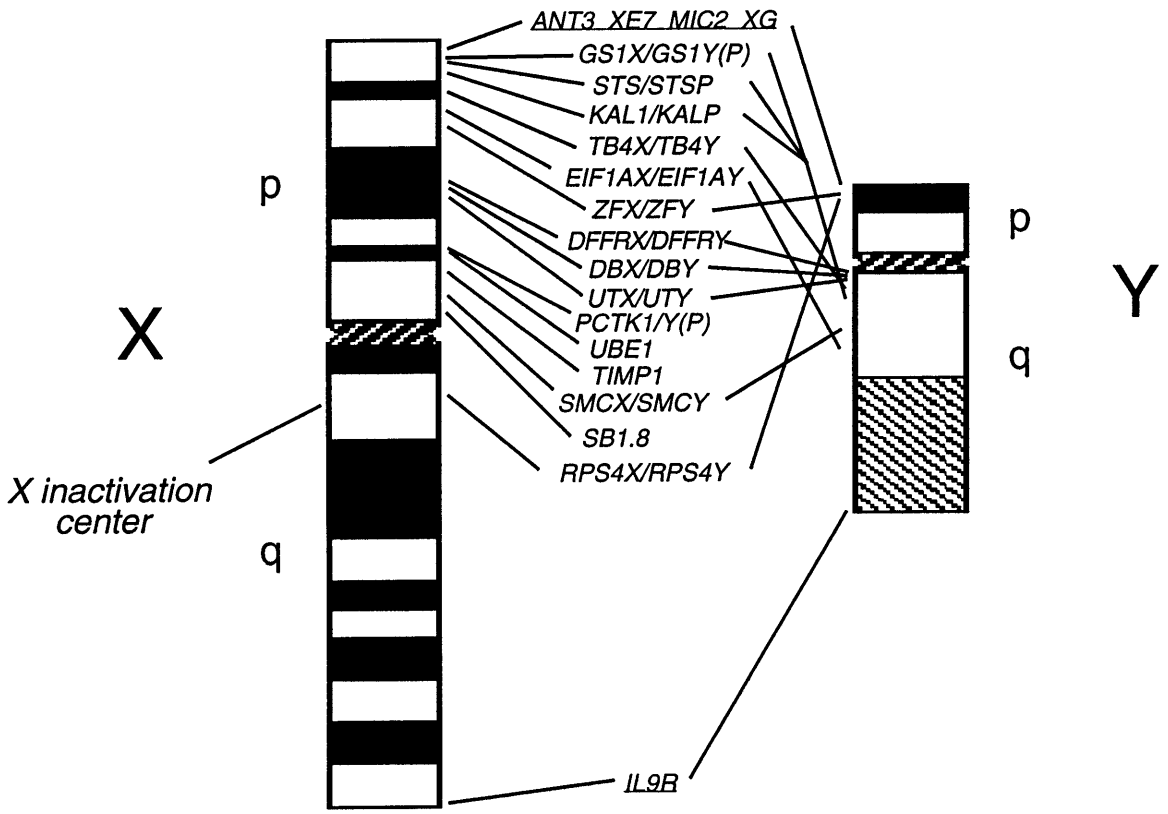


Figure 1.1 Human genes known to escape X inactivation and, when they exist, their Y homologs. Pseudoautosomal genes are underlined.

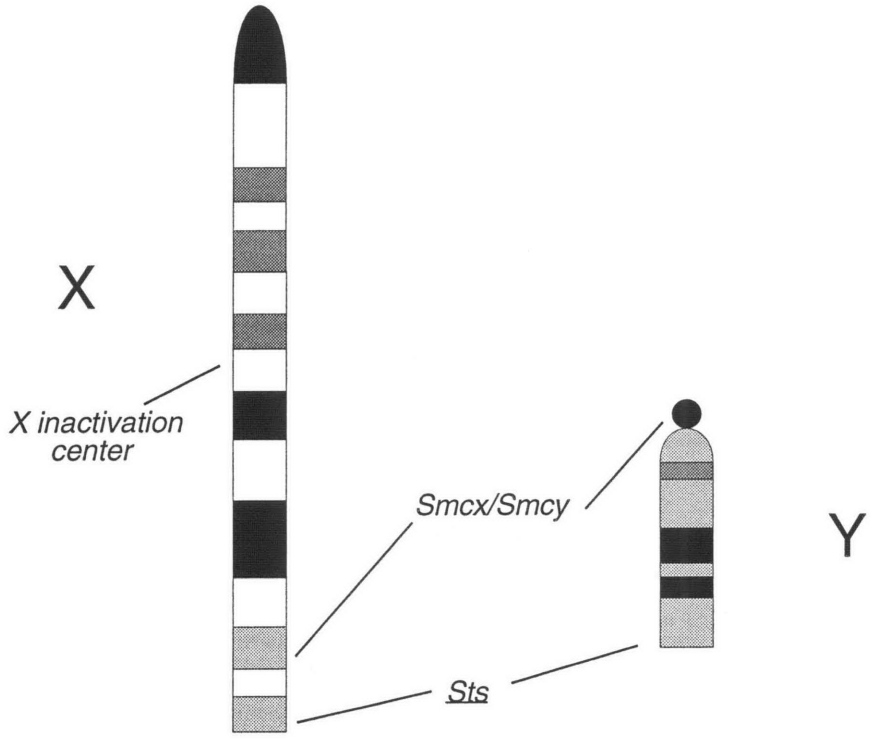


Figure 1.2 Mouse genes known to escape X inactivation and their Y homologs. The pseudoautosomal gene is underlined.

of the X and Y chromosomes. (Two PARs exist, one at the end of each arm, in the acrocentric human sex chromosomes; one PAR has been identified in the mouse, which has telocentric chromosomes.)

In the case of mammals, *SRY* on the Y chromosome triggers male sex differentiation. The emergence of *SRY* seems likely to have been the event that defined the mammalian sex chromosomes. Does a homolog exist on the X that may have descended from the same ancestral gene as *SRY*, a hint of the X and Y's shared autosomal past? Over 20 *SOX* genes—*SRY*-like HMG box-containing genes—are known. Many appear to function in embryonic development. One gene from this family on the X chromosome, *SOX3*, has been suggested as the sister gene to *SRY*; among the *SOX* genes, this gene on the X may be the most closely related to *SRY* in sequence (Graves, 1995). Admittedly, trying to compare a gene's sequence with *SRY* sequence is a precarious business since *SRY* evolves uncommonly quickly. Homologs of *SRY* in different species are unusually divergent (Pamilo and O'Neill, 1997).

Y degeneration and X dosage compensation are logically linked processes. Almost all genes on the chromosome pair that happens to acquire a sex-determining locus will have nothing to do with sex determination and should be evenly expressed in the sexes. If one of these genes decays on the Y, the male will be left with only one, the X-linked, copy of the gene; that gene will be expressed at the same level in females as in males if the female inactivates one X-linked copy. In female mammals, one X chromosome is randomly inactivated to compensate for the loss of Y genes in males. One X chromosome in each cell is stochastically inactivated early in development, in the blastocyst, and thereafter the pattern of inactivation is stably inherited in the somatic cells. But the "inactive" X is not entirely transcriptionally inert. A few exceptional genes exist which are expressed from the inactive X (X_i) as well as from the active X (X_a). Sensibly enough, those genes that escape X

inactivation tend to have retained conserved homologs on the Y chromosome, ensuring that both females and males can express two copies of the gene in any cell.

Escape from X inactivation and conservation of Y homologs is more common in humans than in mice (figures 1.1 and 1.2), the mouse being the one other mammalian species examined with comparable thoroughness at the molecular level. Expression from human and mouse X chromosomes can be fairly compared because of the conservation of the array of genes on the X chromosomes of the two mammals, (except within the pseudoautosomal regions). Other chromosomes have been shuffled during the millions of years of evolution that separate humans and mice though short stretches of synteny persist. The collection of genes on the X chromosomes, however, has remained constant (Copeland et al., 1993). Genes on the X have been shuffled in order, but the same overall set of genes make up the X chromosomes among placental mammals. This constancy reflects the importance of maintaining even gene dosage: once a set of genes has come under some dosage compensation scheme, that set of genes cannot be easily fragmented (Ohno, 1967).

While more and more genes have been examined on the human and mouse X chromosomes, few have been found that evade X inactivation in mice. Only two genes are known to escape X inactivation in mice, one pseudoautosomal gene, *Sts*, and one gene in the sex-linked part of the mouse X, *Smcx* (figure 1.2). Escape from X inactivation appears to be much more common in humans where nineteen genes (four in the pseudoautosomal region and fifteen in the sex-linked region) are known that are not subject to X inactivation (figure 1.1) (Brown et al., 1997; Disteche, 1995).

What we know of the phenotypes associated with monosomy X in humans and mice corroborates the conclusion that the X is more thoroughly inactivated in female mice than in female humans. XO mice are phenotypically nearly normal, difficult to distinguish from their XX sisters. Close examination was required to reveal some growth retardation and the reduction of reproductive lifespan in XO mice

(Burgoyne and Baker, 1981; Burgoyne et al., 1983). (Interestingly, the severity of phenotype characterizing an XO mice depends on the parental origin of the one X chromosome, reflecting, apparently, the fact that the paternal X chromosome is preferentially inactivated in extraembryonic tissue.) In contrast to the nearly normal XO mice, XO humans have Turner Syndrome—a syndrome that is estimated to kill 99% of fetuses with the karyotype *in utero* and leaves the individuals who survive somatically marked and with dysgenetic gonads (Hook and Warburton, 1983). The characteristics of Turner Syndrome evidently reflect the need for the additional dosage of genes provided by the X_i or the Y chromosome.

The difference in gene expression between human and mouse sex chromosomes suggests a question: which of the two dosage strategies—expressing one copy of sex chromosome genes per cell or two copies per cell—represents a more primitive and which a more derived state? Are the exceptional genes, the genes that escape X inactivation and have Y homologs, accidents, eccentric cases, or do they reflect some evolutionary history? Can X inactivation, a seemingly sweeping process, evolve at the level of individual genes? This thesis describes case studies of a set of mammalian sex chromosome genes, studies to determine what expression strategies were used ancestrally, what changes in expression occurred in the course of evolution, and when one dosage strategy may have replaced another.

Approaches for Studying the Evolution of X Inactivation

A principal challenge in an evolutionary study of X inactivation is developing assays that make mammalian species besides the ones with long histories as molecular biology models—that is, humans and mice—accessible for analysis. Short of directly studying a character like X inactivation in ancient animals, the only way we can infer any evolutionary history is to examine a character in a range of extant species and extrapolate primitive states and transitions.

Direct examination of the X inactivation state of a gene requires observing whether one or two alleles are expressed in individual cells, or at least in clonal lineages of cells—not a trivial matter. Historically, X inactivation has been examined using tools specialized for individual species or even for individual genes.

Originally, X inactivation was observed as mosaicism in females heterozygous for certain readily observed phenotypes. In particular, related skin cells with the same X inactivation pattern tend to remain in large, intact, and therefore easily noticed, patches; patchy phenotypes can be observed if a female is heterozygous for some characteristic that appears in the skin, like hair color in mice, which first brought X inactivation to attention (Lyon, 1961), or cats (e.g. the famous calico cats) or like the patchy absence of sweat glands in women with sex-linked anhidrotic ectodermal displasia (Reed et al., 1970). Few genes however give such convenient markers of their activity.

The genes encoding some enzymes, like *G6PD* and *HPRT*, have been shown to be X inactivated by culturing cells from human females with X chromosomes encoding different isozymes of the gene under study. Biochemical tests show that a clonal population of cells from these heterozygous individuals expresses only one variant of the enzyme (Migeon, 1983). The use of this sort of test requires that biochemical tests be available to distinguish forms of an X-encoded enzyme, a rare boon.

Escape from X inactivation has been detected in humans by two principal techniques. RNA levels can be compared convincingly where individuals with X chromosome aneuploidies are available. For instance, transcript levels in XO and XY individuals (with single X chromosomes) can be compared to transcript levels in XX individuals and to transcript levels in rare XXX or even XXXX persons (Schneider-Gädicke et al., 1989). Usually transcript levels are measured in cultured cell lines derived from such persons. If a gene is subject to X inactivation, transcript levels will

be the same in all individuals, but if a gene escapes X inactivation, transcript levels will be proportional to the number of X chromosomes in an individual, e.g. four times greater in an XXXX compared to an XY person. The second method often used to observe escape from X inactivation in humans is examination of expression in hybrid cell lines. Rodent-human cell lines can be created that segregate inactive and active human X chromosomes (Mohandas et al., 1980). Once the two X chromosomes are separated from each other, clones can be expanded and expression analyzed qualitatively: genes that are X inactivated will only be expressed in cell lines with the active X, while genes that escape inactivation will be expressed in lines that contain either the human X_i or X_a. Both of these techniques demand specialized tools—access to individuals with aneuploidies or hybrid cell lines—that are unique to humans. In addition, these methods require that the gene under study be expressed in the cell type from which the cell lines used in these experiments are derived.

Observations of escape from X inactivation in mice have relied on another, also specialized tool, the existence of X chromosome to autosome translocations, the classic case called Searle's translocation, which lead to exclusive inactivation of either the paternal or maternal X chromosome (McMahon and Monk, 1983; Takagi, 1980). In a Searle's translocation mouse, a balanced translocation exists between the X chromosome and chromosome 16. The intact X chromosome tends to be inactivated in all the surviving cells of such mice. If the X inactivation center-carrying portion of the X translocated to chromosome 16 were inactivated, dosage problems would result: only one copy of many chromosome 16 genes would be expressed in each cell while copies of X chromosome genes (from the portion of the X not contiguous with the X inactivation center in these cells) would be overexpressed. With one X consistently inactivated in mice with the balanced translocation, telling whether one or two alleles are active in each cell does not require the examination of separated cells. Exploiting such translocations to study X inactivation requires the creation of mice with two

distinct alleles of an X chromosomal gene; this requirement does not usually present a problem in mice where many and diverged lab strains can be chosen for mating to each other.

Directly assaying X inactivation in any other placental mammal would demand the development of similarly specialized tools in other species or, at the least, would require finding polymorphic alleles gene by gene, species by species, then access to heterozygous individuals whose tissues could be disaggregated. Expression of the polymorphic alleles would have to be examined in separated cells or in clones, if the cells could be cultured. This would be an arduous and tricky process, necessarily unique for every gene and every species examined thus.

In theory, fluorescence in situ hybridization against RNA (that is, RNA FISH) could be used to see if an X gene's transcripts can be observed off only one or off of two chromosomes in a nucleus. This method has been successfully applied once to X inactivation: *UBE1*, which was already known to escape X inactivation, was shown to be expressed from two sites in the human nucleus (Carrel et al., 1996). Despite repeated efforts this method has not worked against *RPS4X*. Apparently, the method cannot be extended reliably to study gene expression though it has the advantage of not requiring the existence of allelic variations to observe expression from one or two alleles.

The complexity and variety of the techniques used to examine X inactivation go to show the difficulty of finding a standard, transferable technique to observe whether one or two alleles of an X-linked gene are expressed in individual cells. Directly gauging X inactivation in diverse mammals does not seem practical. But, fortunately, a convenient surrogate assay for X inactivation exists in placental mammals: 5' CpG island methylation.

About half of all mammalian genes, including all known ubiquitously expressed genes, have CpG island type promoters, a feature of vertebrate genomes (Gardiner-

Garden and Frommer, 1987). CpG islands are typically about a kilobase in length and overlap with the associated genes' transcription start sites. CpG islands are high in G and C nucleotide content and exhibit no diminution in frequency of the dinucleotide CpG. Elsewhere in the vertebrate genome, CpGs are methylated in the germ line and prone to mutate by deamination to TpG (and also, after a round of replication, to CpA), leading to underrepresentation of CpGs. In CpG islands, where CpG dinucleotides are apparently protected from methylation (at least in the germ line), CpG dinucleotides remain at levels similar to the observed levels of GpC dinucleotides. In placental mammals several kinds of experiments have demonstrated that a 5' CpG island of an X-linked gene will be methylated if the gene is X inactivated and not methylated if the gene remains active.

The correlation between X inactivation of a gene and methylation of its 5' CpG island has been illustrated in the following ways. The 5' CpG islands of X inactivated genes like *HPRT* (Yen et al., 1984) and *PGK-1* (Pfeifer et al., 1990) have been shown to be methylated on the inactive X and not methylated on the active X. In one series of experiments, 32 randomly selected CpG islands on the long arm of the human X_i were shown to be hypermethylated while the sister islands on the X_a were found to be uniformly hypomethylated (Tribioli et al., 1992). Also, pseudoautosomal genes on the short arm of the X, which are not X inactivated, exhibit no CpG island methylation on either allele (Goodfellow et al., 1988). In rare cases where normally silent genes on the X_i become reactivated, for example in the chorionic villi or in aged individuals, 5' CpG islands associated with the reactivated genes have been found to be no longer methylated (Migeon et al., 1986; Wareham et al., 1987). In another, perhaps more oblique, demonstration that CpG island methylation correlates with X inactivation, inactive X chromosomes in cells from marsupials, a lineage of mammals that lacks the mechanism of CpG island methylation of X_i genes (Kaslow and Migeon, 1987),

quickly become reactivated when cultured while methylated eutherian Xi chromosomes remain stably silent in tissue culture (Migeon et al., 1989).

In comparison with direct transcriptional measures of X inactivation, using CpG island methylation as a surrogate assay for X inactivation could have several advantages in an evolutionary study. To affirm the efficacy of the strategy, I have demonstrated that CpG island methylation serves as a general marker of X inactivation in placental (eutherian) mammals and is not a trick unique to one or two species (see Chapter 5). As a technique, relying on assays of CpG island methylation has the advantage that it demands access simply to DNA—a commodity easier to obtain and store for a wide range of species than RNA or live tissue. Further, if a gene's 5' region is sufficiently conserved at the level of nucleotide sequence, one methylation assay for a gene may sometimes be applied to a wide range of mammalian species. The PCR-based methylation assays presented throughout this text have the following merits: they do not demand complicated or abundant material from various species, and the results obtained are easily interpretable. Figure 1.3 outlines the basic experimental design: the templates used for PCR are genomic DNA digested with restriction enzymes (REs), among them methylation-sensitive REs,

Figure 1.3 Basic design of methylation assays. PCR primers are designed to flank one or more methylation sensitive restriction endonuclease sites. Genomic DNA is digested with a restriction enzyme before being used as a template for PCR. Female and, wherever available, male DNA was digested with a control enzyme, HindIII, that does not cut any sites between the selected PCR primers; with HpaII or HhaI (latter not drawn), which will not cut if its recognition sites are methylated; and, when HpaII was used, with MspI, an isoschizomer of HpaII that is not sensitive to methylation. If digestion by HpaII or HhaI is complete, PCR product will be observed only if the template was not methylated at the enzyme's recognition sites.

Eutherian Ordinal Phylogeny

Mammals have been the dominant mammals in terrestrial ecosystems for the past 70 million years. The Class Mammalia is thought to have derived from members of the reptilian order Therapsida, small, active carnivores in their prime on Earth about 250 Ma. As a class, mammals are highly active animals, able to be active and adaptable owing to attributes like a four-chambered heart and efficient double circulation, anucleated erythrocytes, the diaphragm, and endothermy. Insulating hair correlates with endothermy. Perhaps mammals' most prominent features, though, are reproductive, marked by the high investment in the young by mothers who bear live young (in all but monotremes) and nourish them with milk secreted from mammary glands.

Mammals have evolved a tremendous variety of forms and habits. Individual animals range in size from a few grams to dozens of tons and live in practically every Earthly habitat. About 4000 extant species of mammals have been classified into about 20 orders with about 120 families. Traditionally, classification has relied on comparative anatomy, with paleontology supplementing the study of living species. In trying to determine higher relationships among mammals, the emphasis has been on grouping species into orders. Efforts to untangle the branching order among major eutherian lineages have tended to be controversial, and the prevailing view of eutherian evolution has been a bush-like radiation occurring between 70 and 100 Ma (Gregory, 1910; Simpson, 1945). To be conservative, the mammalian family trees shown in subsequent chapters (see figures 2.5, 3.4, 4.4, 5.4, 6.1, and A.3) show eutherian orders emanating like the spokes of a wheel from one point, with no attempt to discriminate the branching sequence of the orders.

More recently, molecular approaches have been used to try to reconstruct mammalian phylogeny. Immunological comparisons began to be used around 1970

(Sarich, 1969; Shoshani, 1986). In the mid 1970s, protein sequences began to be compared. But the amount of analysis has exploded since DNA sequences have become accessible. The new wealth of data has helped clarify some points in eutherian systematics, but it has also raised some controversies that have yet to be resolved. Altogether, no consensus has yet been attained among those who study eutherian phylogenetics about the exact sequence of eutherian radiation.

Nucleotide sequence analysis has been used to show that some morphogenetically well-recognized orders may not be monophyletic (or members of a clade, a group that shares the same ancestor, an ancestor not shared by those outside the group). For example, comparison of protein and mitochondrial DNA sequences suggested that all the animals considered rodents probably do not constitute a clade. Specifically, hystricomorphs (guinea pigs, porcupines)* appear not to be related to myomorphs (mice, rats, hamsters, lemmings) or sciuromorphs (squirrels) more than to other eutherian orders and so appear to deserve to be classified as an order unto themselves (Graur et al., 1992; Ma et al., 1993). Understanding the fine points of rodent relationships help in interpreting the evolution of *ZFX* and *ZFY* (Chapter 2). The group including both the myomorphs and the sciuromorphs is called sciurognathi. In keeping with the unsettled nature of many phylogenetic questions, it is not yet clear that the hystricomorphs are a true clade. Guinea pigs (caviomorphs) have been used mostly in comparisons of hystricomorphs with sciurognathi.

While some groups seem to be falling apart, other orders that were classified separately are coalescing under molecular examination. In particular, protein sequence (e.g. comparison of 39 mammalian ribonuclease sequences (Beintema et al., 1986)) and mitochondrial sequence data (Arnason et al., 1991; Graur and Higgins, 1994) place cetaceans (whales) not as a separate order but rightly nestled among artiodactyls (cattle, pigs, llamas). Comparison of DNA sequences over 1 kb long in the 5' end of

* Species indicated in paranthesis after a family name, here and henceforth, are meant to provide familiar examples rather than a comprehensive list.

exon 1 of the IRBP (interphotoreceptor retinoid binding protein) gene confirm that artiodactyls and cetaceans constitute a clade (Stanhope et al., 1996). Graur and Higgins declared, based on mitochondrial DNA sequence analysis, that cetaceans are not only intimately related to the artiodactyls—they *are* artiodactyls (Graur and Higgins, 1994); they estimated that cetaceans diverged from ruminants (cattle, sheep, pigs, goats, deer) some 50 Ma, after their lineage diverged from the other suborders of artiodactyls, suiformes (hippopotamus, peccary) and tylopods (llama, alpaca, guanaco). The high confidence of this conclusion helps in interpreting *SMCX* evolution (Chapter 4).

Deducing the branching sequence of the various eutherian orders remains something of a holy grail among taxonomists. Multifurcated trees, as I draw conservatively in the remainder of this work, are anathema to taxonomists. Since speciation is a binary process, all nodes of a true phylogenetic tree should be bifurcated. A multifurcation means that a binary resolution could not be attained. As Graur points out, finding the one, true bifurcating tree that would account for relationships among all the eutherian orders requires the proper identification of one cladogram from approximately 10^{19} possible phylogenetic trees (Graur, 1993).

Rather than try to reconstruct the relationships of all the eutherian orders at once, experimental studies have tended to concentrate on the relationships of just a few orders at a time, partly to simplify analysis and partly because sequence data tend to be richer for some species than for others. For instance, a greater wealth of sequence information has been accumulated for the primates, rodents, lagomorphs (rabbits, hares, pikas), artiodactyls, and carnivores (dogs, cats, bears, skunks, seals, walruses) than other orders, so conclusions about relationships among these orders are likely to be more reliable than when discussion turns to less well characterized orders, like edentates and insectivores.

One prominent point of contention has been whether the Cohort Glires, including rodents and lagomorphs, is real. Anatomic studies—examination of skull structure, ankle joints, fetal membranes, and tooth development (Novacek et al., 1988)—promoted this classification. But molecular studies examining nucleotide and amino acid sequences of many genes have contradicted the grouping. Instead, surprisingly, lagomorphs may be closely related to primates, and rodents may have branched far earlier (Easteal, 1990; Graur et al., 1996). And, as discussed earlier, different groups of the animals known as rodents have probably diverged at different times in the course of eutherian divergence.

Several studies have placed carnivores close to the artiodactyl/cetacean clade and suggest that perissodactyls (horses, zebras, rhinoceroses, tapirs) may be next closest to those three groups. Full comparisons of mitochondrial DNA sequences (greater than 15 kb in each species) argued for this view (Krettek et al., 1995). Comparisons of more than 1 kb of IRBP sequence reiterated it (Stanhope et al., 1996).

Rodents—the sciurognathi (myomorphs and sciuromorphs) as well as the hystricomorphs, newly elevated to ordinal status—may be among the eutherian lineages that diverged first from the eutherian line. Comparisons of the coding regions of many genes and of mitochondrial DNA sequences have suggested that rodents are outgroups to, at least, lagomorphs, primates, artiodactyls, and carnivores (Cao et al., 1994; Easteal, 1990; Li et al., 1990).

Insectivores (represented by hedgehog) may be even more distant relatives among placental mammals. When its full 17.4 kb mitochondrial DNA sequence was compared to the sequences known in other mammals, hedgehog appeared to be an outgroup to primates, rodents, artiodactyls, cetaceans, carnivores, and perissodactyls (Arnason et al., 1991; Krettek et al., 1995).

But when asked to suggest the eutherian order most likely to have diverged first, edentates (anteaters, armadillos, sloths) are often cited as the most distant relatives to the other placental mammals. Edentates have not been as extensively studied as the representatives of many other orders (and may be ripe for disputations about monophyly), but examination of IRBP sequences, 1.2 kb from 25 species representing 17 eutherian orders, places edentates as the first to diverge (Stanhope et al., 1996). Amino acid sequence comparisons of 39 mammalian ribonucleases had suggested edentates as the first to diverge a decade before (Beintema et al., 1986).

I have found no reports suggesting that primates diverged first among the eutherian orders. On the other hand, no one has firmly associated the primates with other orders, as for example the carnivores are associated with artiodactyls and cetaceans. The point of divergence of primates during eutherian radiation is an unsettled question, relevant in the context of *RPS4* evolution (Chapter 3).

In figure 1.4 two phylogenetic trees are shown, a conservative tree that does not attempt to discriminate the branching sequence of the eutherian orders, showing all the represented orders emanating from one multifurcated point. A second tree shows what a fully bifurcated cladogram representing the same species might look like based on accumulated sequence data; many points of bifurcation in the tree are undoubtedly true, but it is likely that the presented tree is not exactly right. After all, millions of possible fully bifurcated cladograms of 11 orders can be drawn.

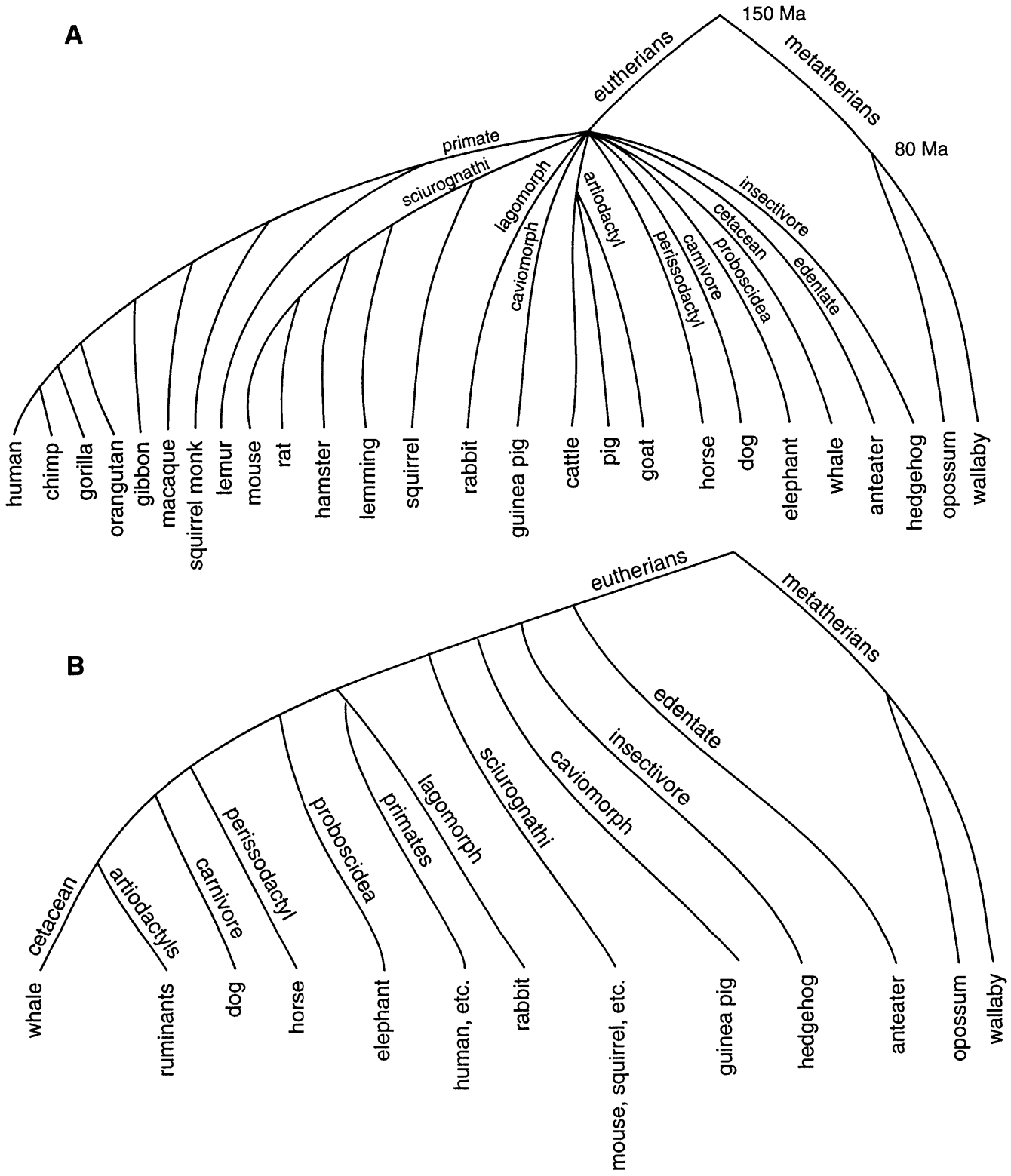


Figure 1.4 Two representations of mammalian phylogeny that include all the species examined in this text. (**A**) A conservative representation showing all the eutherian orders emanating from a single point. This picture is repeated when data are displayed in phylogenetic context in the following chapters. (**B**) A more speculative depiction where only bifurcating nodes are shown.

References

Arnason, U., Gullberg, A., and Widegren, B. (1991). The complete nucleotide sequence of the mitochondrial DNA of the fin whale, Balaenoptera physalus. *J Mol Evolution 33*, 556-568.

Beintema, J. J., Fitch, W. M., and Carsana, A. (1986). Molecular evolution of pancreatic-type ribonucleases. *Mol Biol Evol 3*, 262-275.

Brown, C. J., Carrel, L., and Willard, H. F. (1997). Expression of genes from the human active and inactive X chromosomes. *Am J Hum Genet 60*, 1333-1343.

Burgoyne, P. S., and Baker, T. G. (1981). Oocyte depletion in XO mice and their XX sibs from 12 to 200 days *post partum*. *J Reprod Fertil 61*, 207-212.

Burgoyne, P. S., Evans, E. P., and Holland, K. (1983). XO monosomy is associated with reduced birthweight and lowered weight gain in the mouse. *J Reprod Fertil 68*, 381-385.

Cao, Y., Adachi, J., and Hasegawa, M. (1994). Eutherian phylogeny as inferred from mitochondrial DNA sequence data. *Jpn J Genet 69*, 455-472.

Carrel, L., Clemson, C. M., Dunn, J. M., Miller, A. P., Hunt, P. A., Lawrence, J. B., and Willard, H. F. (1996). X inactivation analysis and DNA methylation studies of the ubiquitin activating enzyme E1 and PCTAIRE-1 genes in human and mouse. *Hum Molec Genet 5*, 391-401.

Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., Steen, R. G., and et, a. (1993). A genetic

linkage map of the mouse: current applications and future prospects. *Science 262*, 57-66.

Disteche, C. M. (1995). Escape from X inactivation in human and mouse. *Trends Genet 11*, 17-22.

Easteal, S. (1990). The pattern of mammalian evolution and the relative rate of molecular evolution. *Genetics 124*, 165-173.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. *Journal of Molecular Biology 196*, 261-282.

Goodfellow, P. J., Mondello, C., Darling, S. M., Pym, B., Little, P., and Goodfellow, P. N. (1988). Absence of methylation of a CpG-rich region at the 5' end of the MIC2 gene on the active X, the inactive X, and the Y chromosome. *Proc Natl Acad Sci USA 85*, 5605-5609.

Graur, D. (1993). Towards a molecular resolution of the ordinal phylogeny of the eutherian mammals. *FEBS 325*, 152-159.

Graur, D., Duret, L., and Gouy, M. (1996). Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). *Nature 379*, 333-335.

Graur, D., Hide, W. A., Zharkikh, A., and Li, W. H. (1992). The biochemical phylogeny of guinea-pigs and gundis, and the paraphyly of the order rodentia. *Comp Biochem Physiol [B] 101*, 495-498.

Graur, D., and Higgins, D. G. (1994). Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. *Mol Biol Evol 11*, 357-364.

Graves, J. A. (1995). The evolution of mammalian sex chromosomes and the origin of sex determining genes. *Philos Trans R Soc Lond B Biol Sci 350*, 305-312.

Gregory, W. K. (1910). The orders of mammals. *Bull Am Mus Nat Hist 27*, 1-524.

Hook, E. B., and Warburton, D. (1983). The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism. *Hum Genet 64*, 24-27.

Kaslow, D. C., and Migeon, B. R. (1987). DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation. *Proc Natl Acad Sci USA 84*, 6210-6214.

Krettek, A., Gullberg, A., and Arnason, U. (1995). Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, Erinaceus europaeus, and the phylogenetic position of the Lipotyphla. *J Mol Evolution 41*, 952-957.

Li, W. H., Gouy, M., Sharp, P. M., O'H Uigin, C., and Yang, Y. W. (1990). Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. *Proc Natl Acad Sci USA 87*, 6703-6707.

Lyon, M. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus L*). *Nature 190*, 372-373.

Ma, D. P., Zharkikh, A., Graur, D., Vande Berg, J. L., and Li, W. H. (1993). Structure and evolution of opossum, guinea pig, and porcupine cytochrome b genes. *J Mol Evolution 36*, 327-334.

McMahon, A., and Monk, M. (1983). X-chromosome activity in female mouse embryos heterozygous for Pgk-1 and Searle's translocation, T(X; 16) 16H. *Genetic Research 41*, 69-83.

Migeon, B. R. (1983). Glucose-6-phosphate dehydrogenase as a probe for the study of X-chromosome inactivation in human females. *Isozymes Curr Top Biol Med Res 9*, 189-200.

Migeon, B. R., de Beur, S. J., and Axelman, J. (1989). Frequent derepression of G6PD and HPRT on the marsupial inactive X chromosome associated with cell proliferation in vitro. *Experimental Cell Research 182*, 597-609.

Migeon, B. R., Schmidt, M., Axelman, J., and Cullen, C. R. (1986). Complete reactivation of X chromosomes from human chorionic villi with a switch to early DNA replication. *Proc Natl Acad Sci USA 83*, 2182-2186.

Mohandas, T., Sparkes, R. S., Hellkuhl, B., Grzeschik, K. H., and Shapiro, L. J. (1980). Expression of an X-linked gene from an inactive human X chromosome in mouse-human hybrid cells: further evidence for the noninactivation of the steroid sulfatase locus in man. *Proc Natl Acad Sci USA 77*, 6759-6763.

Novacek, M. J., Wyss, A. R., and McKenna, M. C. (1988). *The major groups of eutherian mammals.* In The Phylogeny and Classification of the Tetrapods, M. J. Benton, ed. (Oxford: Clarendon), pp. 31-71.

Ohno, S. (1967). *Sex Chromosomes and Sex-Linked genes* (Berlin: Springer Verlag).

Pamilo, P., and O'Neill, R. J. (1997). Evolution of the Sry genes. *Mol Biol Evol 14*, 49-55.

Pfeifer, G. P., Tanguay, R. L., Steigerwald, S. D., and Riggs, A. D. (1990). In vivo footprint and methylation analysis by PCR-aided genomic sequencing: Comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. *Genes Dev 4*, 1277-1287.

Reed, W. B., Lopez, D. A., and Landing, B. (1970). Clinical spectrum of anhidrotic ectodermal dysplasia. *Arch Dermatol 102*, 134-143.

Sarich, V. M. (1969). Pinniped origins and the rate of evolution of carnivore albumins. *Syst Zool 18*, 286-295.

Schneider-Gädicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R., and Page, D. C. (1989). *ZFX* has a gene structure similar to *ZFY*, the putative human sex determinant, and escapes X inactivation. *Cell 57*, 1247-1258.

Shoshani, J. (1986). Mammalian phylogeny: comparison of morphological and molecular results. *Mol Biol Evol 3*, 222-242.

Simpson, G. G. (1945). The principles of classification and a classification of mammals. *Bull Am Mus Nat Hist 85*, 1-350.

Stanhope, M. J., Smith, M. R., Waddell, V. G., Porter, C. A., Shivji, M. S., and Goodman, M. (1996). Mammalian evolution and the interphotoreceptor retinoid binding protein (IRBP) gene: convincing evidence for several superordinal clades. *J Mol Evolution 43*, 83-92.

Takagi, N. (1980). Primary and secondary nonrandom X chromosome inactivation in early female mouse embryos carrying Searle's translocation T(X; 16)16H. *Chromosoma 81*, 439-459.

Tribioli, C., Tamanini, F., Patrosso, C., Milanesi, L., Villa, A., Pergolizzi, R., Maestrini, E., Rivella, S., Bione, S., Mancini, M., and et, a. l. (1992). Methylation and sequence analysis around EagI sites: Identification of 28 new CpG islands in XQ24-XQ28. *Nucl Acids Res 20*, 727-733.

Wareham, K. A., Lyon, M. F., Glenister, P. H., and Williams, E. D. (1987). Age related reactivation of an X-linked gene. *Nature 327*, 725-727.

Yen, P. H., Patel, P., Chinault, A. C., Mohandas, T., and Shapiro, L. J. (1984). Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes. *Proc Natl Acad Sci USA 81*, 1759-1763.

Chapter 2. *ZFX/ZFY*: Caught in Transition

Background

Genes common to the X and Y chromosomes were first observed in the mammalian sex chromosomes' pseudoautosomal regions where recombination between the X and the Y enforces sequence identity (Fialkow, 1970; Goodfellow et al., 1984). *ZFX* and *ZFY* were among the first X-Y homologous genes found in the sex-linked portions of the mammalian X and Y chromosomes, where no recombination occurs (Schneider-Gädicke et al., 1989; Shapiro et al., 1979). *ZFX* and *ZFY* encode zinc-finger proteins thought to act as transcriptional activators. Each possesses an amino-terminal acidic domain, a putative nuclear localizing signal, and a carboxy-terminal domain of 13 zinc fingers (Schneider-Gädicke et al., 1989). *ZFX* has been shown to contribute to embryonic cell growth, animal size, and germ cell number in mice (Luoh et al., 1997). Homologs of *ZFX* have been found on the X chromosomes of all placental mammals tested. In humans, *ZFX* and *ZFY* are both expressed ubiquitously and are conserved in coding sequence—in some exons conserved to 98% identity at the nucleotide or amino acid level—while considerably diverged in introns, untranslated exons, and flanking sequences (Schneider-Gädicke et al., 1989; Shimmin et al., 1993), suggesting that the X and Y homologs have experienced similar selection pressures at the protein level and are therefore likely conserved in function.

One portion of the noncoding sequences of *ZFX* and *ZFY* is exceptionally well conserved, an exception to the general observation that these genes are more similar in coding than in noncoding exons. The genes' 5' CpG islands are peculiarly conserved in sequence (Luoh et al., 1995). Between human *ZFX* and *ZFY* the CpG island is about 88% conserved. Between human *ZFX* and mouse *Zfx* the level of conservation is closer to 93%. (Observation of greater conservation between the two X genes compared to the two human genes suggests that sequence similarity in the CpG islands reflects simple

conservation, rather than some kind of gene conversion event within the primate lineage.) Over one remarkable stretch of 165 bp that extends from the first 5' untranslated exon into the first intron, human *ZFX* and mouse *Zfx* are absolutely identical to each other, and the X-linked genes are 95% identical to human *ZFY* (figure 2.1). Figure 2.2 verifies the CpG island character of the 5' regions of *ZFX* genes.

When *ZFX* and *ZFY* were first discovered in humans, Southern blotting experiments revealed the existence of highly conserved Y-chromosomal homologs in several other mammals but not in mice. Two Y homologs of *ZFX* were eventually uncovered in mice, *Zfy1* and *Zfy2* (Mardon et al., 1990; Mardon and Page, 1989), but these genes have substantially diverged at the nucleotide and amino acid levels, accounting for the failure to detect them by high-stringency Southern blotting using human *ZFX*, *ZFY* or mouse *Zfx* probes. For example, in the Zinc-finger region, human *ZFX*, *ZFY*, and mouse *Zfx* are 92% or more conserved among each other at the nucleotide level while the three genes share no more than 82% identity with mouse *Zfy1* and *Zfy2* in the same region. Also unlike the other genes, the two mouse *Zfy* genes employ TATA box rather than CpG island promoters (Zambrowicz et al., 1994). Also, in contrast to the other ubiquitously expressed genes, the mouse Y homologs appear to be expressed solely in the testes, at least in the adult mouse (Nagamine et al., 1989; Zambrowicz et al., 1994). The loss of the 5' CpG island during mouse *Zfy* evolution seems apt given that mammalian genes with CpG islands tend to be more widely expressed than those without CpG island promoters. Finally, in mice, an intact Y chromosome does not complement defects in growth and viability caused by a mutation in *Zfx* (Luoh et al., 1997), again suggesting that the mouse Y-chromosomal genes have a restricted or diverged function compared with their mouse X, human X, and human Y homologs.

When human *ZFX* was discovered, it was found to escape X inactivation by using somatic cell hybrids that separated the inactive and active X chromosomes and by

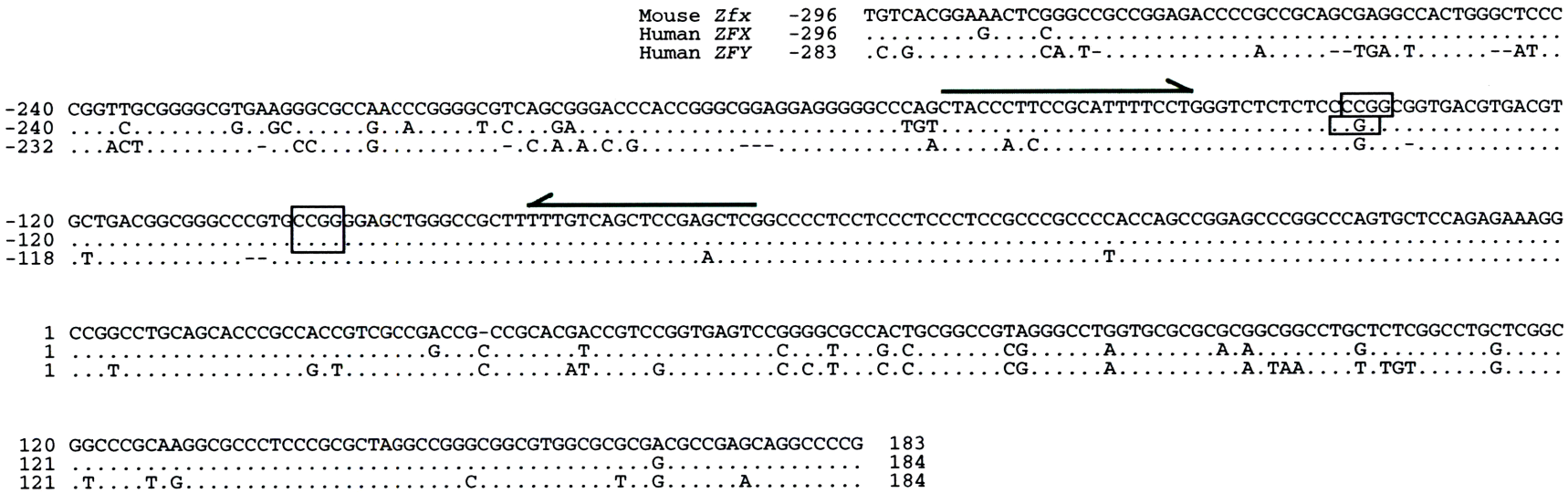


Figure 2.1 Alignment of mouse *Zfx*, human *ZFX*, and human *ZFY* sequences in the 5' CpG island regions. HpaII/MspI (CCGG) sites are shown boxed. PCR primers used in methylation assays are outlined.

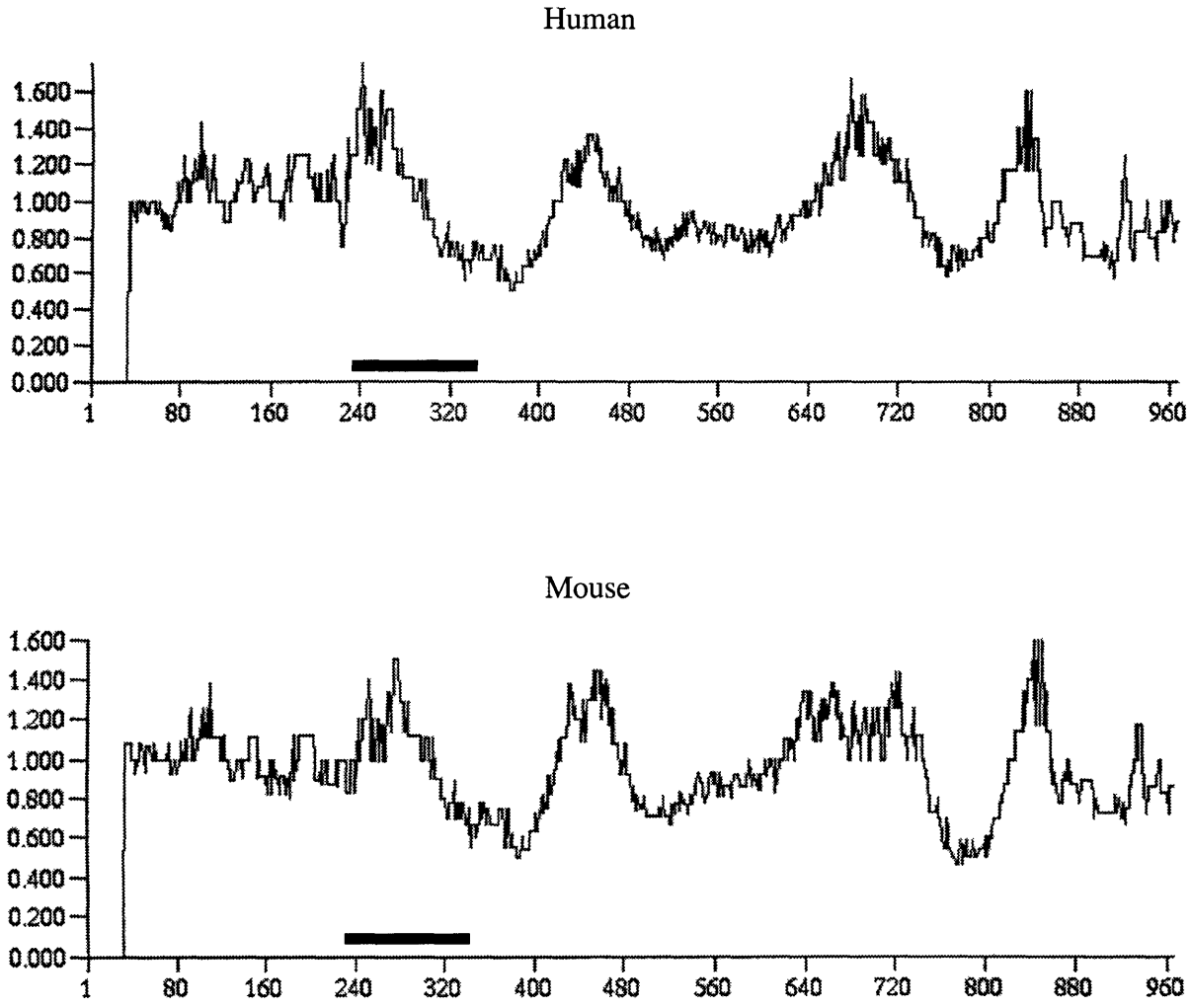


Figure 2.2 CpG to GpC ratios in the 5' genomic regions of human and mouse *ZFX*. In mammalian genomic DNA the CpG to GpC ratio tends to be near 0.2 but in CpG islands approaches 1. The sites where methylation was assessed lie near 250 bp in these graphs. Areas where methylation was tested are marked with bars over the *x* axes. Dinucleotide ratios were calculated in 70 bp windows.

Note that discrepancies are expected at the 5' and 3' ends of these kinds of graphs because dinucleotide counting occurs over windows.

Northern blots that used tissue from individuals with X chromosome aneuploidies (Schneider-Gädicke et al., 1989). When mouse *Zfx* expression was examined by interspecies crosses, the mouse gene was found to be X inactivated (Adler et al., 1991).

The accumulated information from humans and mice about X inactivation and the existence of Y homologs of *ZFX* sorted into something seemingly sensible: *ZFX* is X inactivated in mice and escapes inactivation in humans; a conserved Y homolog of *ZFX* exists in humans but only diverged Y homologs exist in mice. Either way, if a conserved *ZFY* can functionally mimic a *ZFX*, overall gene dosage should be equal in males and females: in mice, only one copy of the gene is expressed per cell no matter what the animal's sex, while in humans, two copies of the gene are expressed per cell no matter an individual's sex.

Which dosage strategy—that demonstrated by humans or that in mice—represents the more primitive, the more original state and which a more derived, more evolved state?

Methods

Sources of genomic DNA. Genomic DNA, used for methylation assays in this and subsequent chapters, was extracted from the following tissue or cell types: spleen or liver in mouse samples; liver in rat, hamster, whale, wallaby, and opossum; kidney or liver in guinea pig; lymphoblastoid cell lines in human; blood cells in all other primates, rabbit, horse, cow, goat, dog, and elephant. Cells were lysed, treated with proteinase K, phenol/chloroform extracted, and the DNA was dialyzed.

Methylation analysis of genomic DNAs. One hundred nanograms of genomic DNA was incubated with 10 units HindIII or HpaII or MspI for 4 hours at 37°C in buffers recommended by the manufacturer (New England Biolabs). These digested genomic

DNAs were then used as template in PCR with primers (CTACCCTTCCGCATTTTCCT and GAGCTCGGAGCTGACAAAAA) chosen from sequences conserved between mouse *Zfx* and human *ZFX* and spanning, in both species, a 105-bp region containing two CCGG sites (figure 2.1). PCR using 100 ng template DNA was carried out in 20 μL of 12.5mM Tris pH 8.2, 50mM KCl, 12.5mM NaCl, 5mM NH_4Cl, 2.5mM $MgCl_2$, and 1mM of each of the two primers. After heating to 100°C for 5 min, the four deoxyribonucleotides (to a final concentration of 0.125mM each) and 2 units Taq polymerase were added. Thirty cycles of 1 min at 94°C, 1 min at 62°C, and 1 min at 72°C were followed by extension for 2 min at 72°C.

PCR products were subjected to electrophoresis in 4% NuSieve agarose (FMC Corporation), 90mM Tris-borate, 2mM EDTA, 0.5 mg/mL ethidium bromide; visualized with UV light. In the case of human and mice, products were also transferred to nylon membrane in preparation for Southern hybridization. The hybridization probe was an oligonucleotide, GGTGACGTGACGTGCTGACG, chosen from sequence within the PCR product that was conserved completely between mouse and human. The oligonucleotide was labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase and allowed to hybridize with the filter overnight at 42°C in 6X SSC, 5X Denhardt's, 0.05% $Na_4P_2O_7$, 100 mg/mL tRNA, and 0.5% SDS. The filter was then washed three times for 20 min each at 42°C in 6X SSC, 0.1% SDS and exposed to X-ray film for two days.

Southern blot analysis of genomic DNAs. Genomic DNAs were prepared from blood, kidney, or liver samples of males and females from various species, digested with EcoRI (at 37° in 50 mM NaCl, 100 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM spermidine), electrophoresed on a 0.75% agarose gel, and transferred (Southern, 1975) to nylon membrane. A 395-bp *Bss*HII genomic fragment from human *ZFY* was labeled with ^{32}P by random-primer synthesis (Feinberg and Vogelstein, 1984) and hybridized overnight to the genomic DNA transfer at 67°C in 1 mM EDTA, 0.5 M $NaPO_4$ pH 7.2, 7% SDS

(sodium dodecyl sulfate) (Church solution). Following hybridization, the transfer membrane was washed three times for 20 min each at 62°C in 0.1X SSC, 0.1% SDS and exposed at -80°C with X-ray film backed with an intensifying screen for one day. Alternatively probe pDP1007, a 1.3 kb genomic fragment that includes the Zinc finger exon of human *ZFY* (Page et al., 1987), was labeled with ^{32}P, and hyridized to the genomic DNA transfer at 65°C in Church solution, then washed and otherwise treated as above.

Results

Knowing the X inactivation status of *ZFX* in humans and mice and that methylation of CpG islands has been found to correlate with X inactivation, I tested the prediction that methylation of CpG islands would parallel differences in X inactivation for a gene that is known to be differentially inactivated between humans and mice. I assayed the methylation status of two consecutive HpaII/MspI (CCGG) sites in the most highly conserved portion of the CpG island. HpaII cleaves only when the CpG dinucleotide it recognizes is unmethylated while MspI cleaves the site regardless of methylation status. Female and male genomic DNAs digested with one of three restrictions endonucleases (HpaII, MspI, or HindIII, which does not cut in the vicinity and was used as a control) were used as alternate templates for PCR reactions with primers flanking the two CCGG sites (figure 2.1). In the case of complete digestion, no PCR products should be observed after MspI digestion, and PCR amplification should only be observed from HpaII-digested template if the sites were methylated.

With DNA from either human or mouse males, little or no PCR amplification was seen after HpaII (or MspI) digestion (figure 2.3). Since males carry only a single,

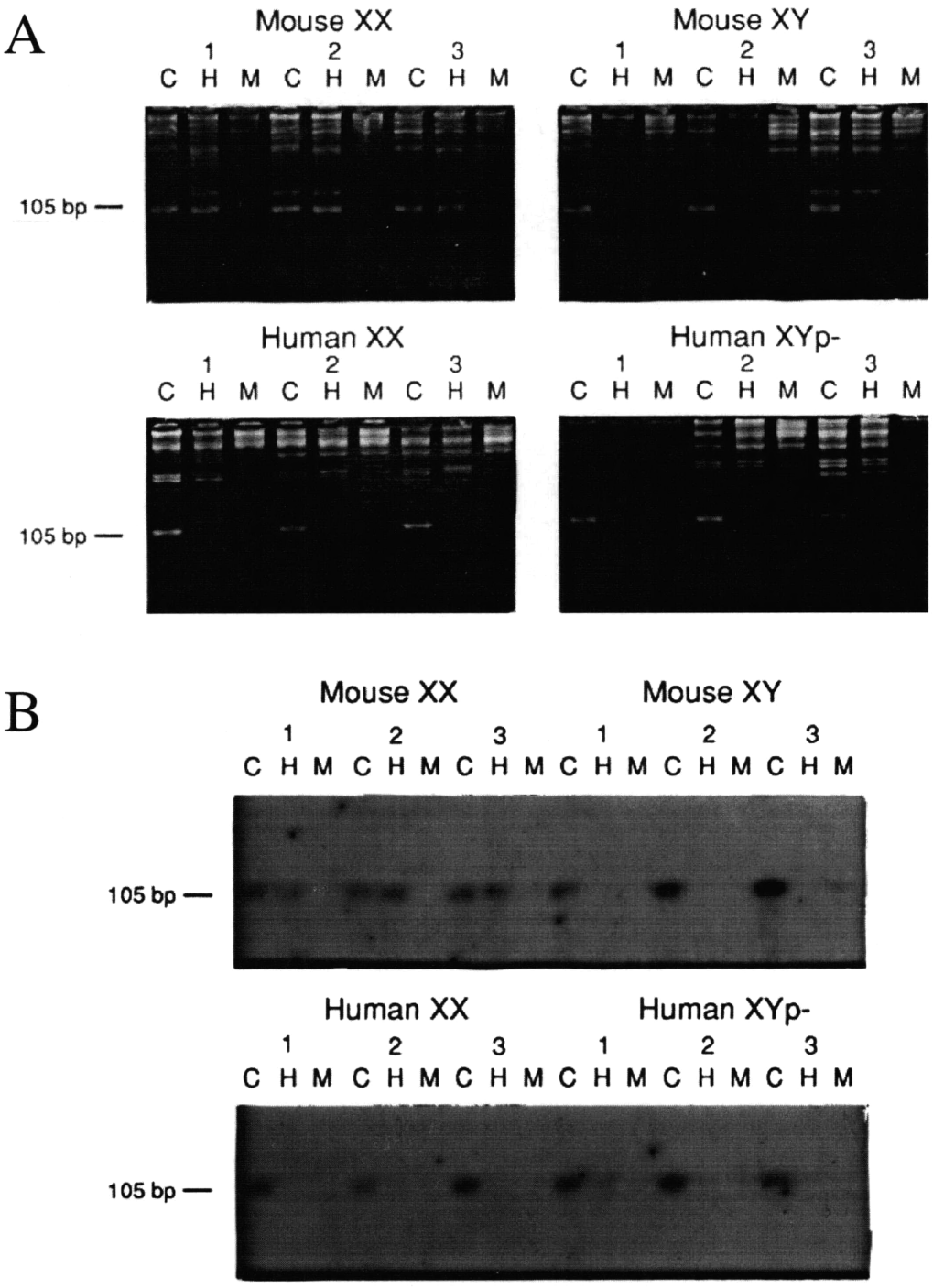


Figure 2.3 Differences in *ZFX* CpG island methylation between humans and mice. Mouse and human genomic DNAs digested with HpaII (H), MspI (M), or HindIII (C, "control") were used as alternate templates in the PCR assay described in figure 2.1. (**A**) PCR products from three individuals of each species and sex chromosome constitution visualized by ethidium bromide/UV staining after agarose gel electrophoresis. Human XYp- individuals lack the *ZFY* gene (and retain *ZFX*) (Cantrell et al, 1989). (**B**) Southern blot autoradiogram of the same gel hybridized with oligonucleotide internal to PCR primers.

active X chromosome, this indicated that the sites tested are unmethylated on active X chromosomes in both mouse and human, as one could have predicted. Using female DNAs, derived from cells with one active and one inactive X chromosome, we observed PCR amplification after HpaII (but not MspI) digestion of mouse female DNA, presumably reflecting methylation of the inactivated *Zfx* allele. But with human female DNAs, no amplification was observed after HpaII digestion, indicating that the *ZFX* CpG island is unmethylated on both the active and inactive X chromosomes. Thus, methylation of the X-linked gene's CpG island reflects the expression status of the gene rather than the inactivation status of the host chromosome.

The extraordinary conservation of the nucleotide sequence of the *ZFX* CpG island allowed this methylation assay to be extended, without modification, to a wide array of mammalian species. I assayed methylation in male and female genomic DNAs from 19 eutherian species representing eleven orders (figures 2.4, 2.5). In no species did male DNA show evidence of methylation of the *ZFX* CpG island, as expected given that the single X chromosome in males is active. In 15 of the 19 eutherian species tested, female DNA also exhibited no methylation, indicating that *ZFX* escapes X inactivation in those species. Methylation in females was found in only four species: mouse, rat, hamster, and lemming, all of which belong to the single taxonomic group, myomorpha. Two marsupial species, where the homolog of *ZFX* is autosomal (Sinclair et al., 1988) and we would expect the CpG island to be unmethylated—as in autosomal CpG islands generally—indeed showed no *ZFX* methylation.

Having observed no *ZFX* methylation in species other than myomorphs, the question remained open whether 5' CpG island methylation is even associated with X inactivation in other eutherians. Chapter 5, describing methylation results for the X-inactivated *ALD* gene, addresses this reservation and shows that methylation is indeed correlated with the presence of an inactive X allele in a broad range of placental mammals. An implicit assumption here and in subsequent chapters is that in all the cell

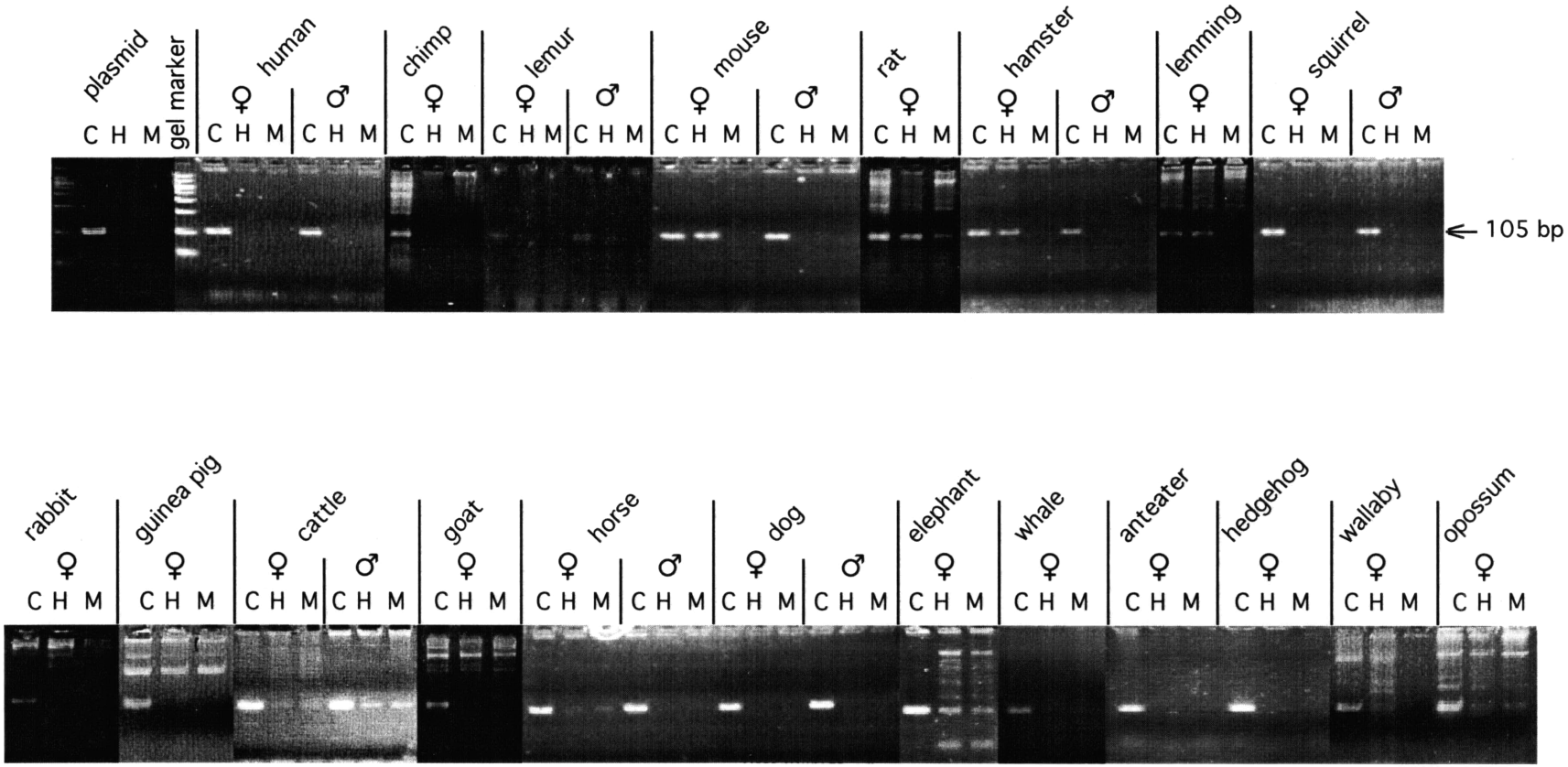


Figure 2.4 *ZFX* methylation results. Female and male genomic DNAs digested with HpaII (H) or MspI (M) or HindIII (C, "control") were used as alternate templates in PCR assay diagrammed in figure 2.1. PCR products were visualized by ethidium bromide/UV staining after agarose gel electrophoresis.

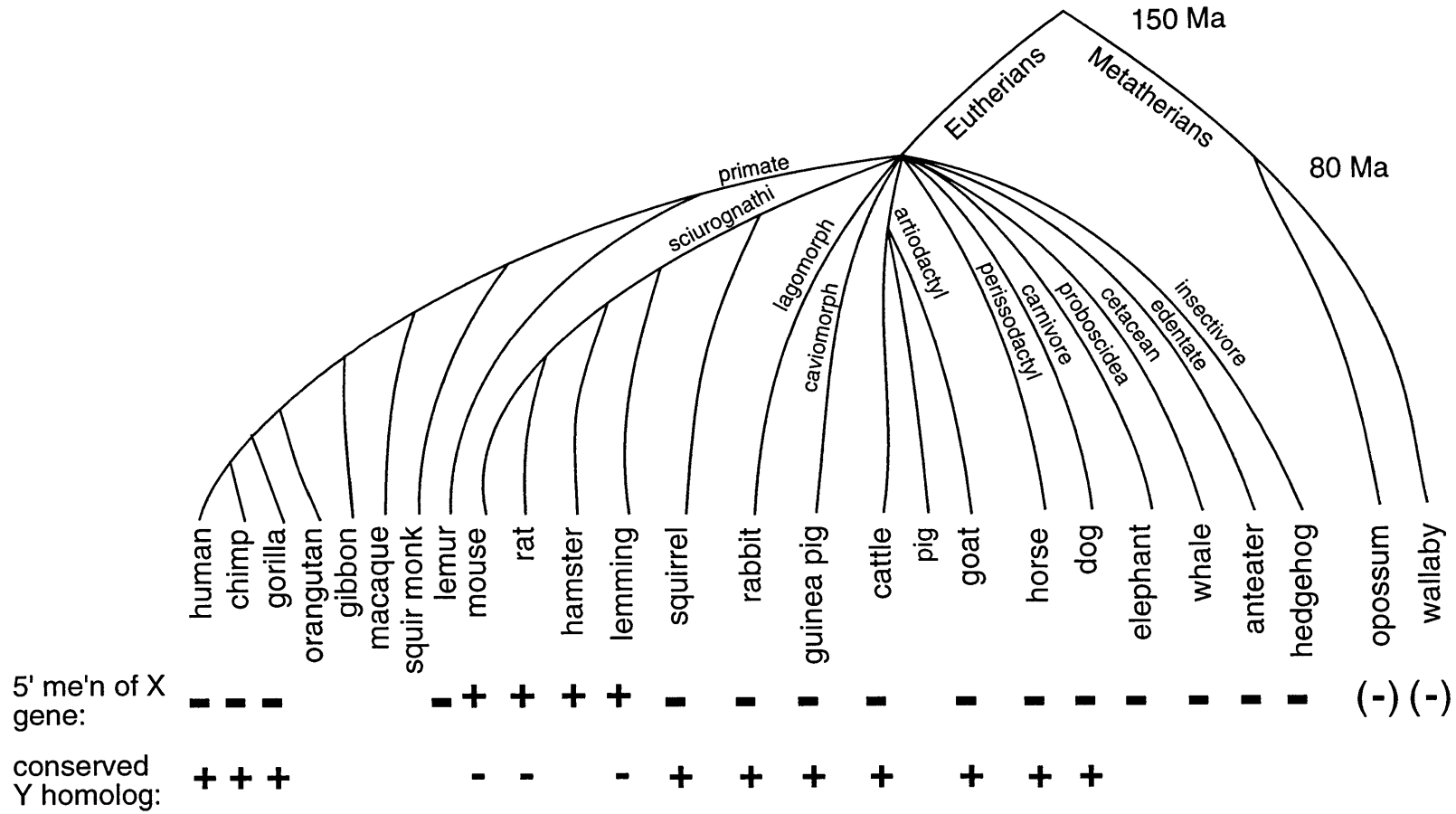


Figure 2.5 Results of *ZFX* methylation and Y homolog studies in mammalian species, arranged phylogenetically. For marsupials, *ZFX* methylation results are shown in paranthesis since the marsupial *ZFX* homolog is autosomal and therefore cannot be subject to X inactivation. All species for which *ZFY* homolog results are reported were tested with two Southern hybridization probes derived from the human *ZFY* gene, a CpG island probe (figure 2.6) and zinc-finger probe pDP1007 (Page et al, 1987, additional primary data not shown).

types tested, whether derived from liver, spleen, blood, kidney, or lymphoblastoid lines, 5' CpG island methylation will be consistent; that is, methylation will not depend on the tissue from which DNA was derived. Indeed, the only tissues where X chromosome methylation patterns have been shown to differ from patterns in other tissues are the placenta and germ cells (Singer-Sam and Riggs, 1993).

The mammalian phylogenetic tree provides a simple way to organize and interpret these results (figure 2.5). The four species showing methylation of the *ZFX* CpG island are monophyletic, all descended from a common ancestor not shared by any of the other 15 eutherian species examined. These results suggest that in the common ancestor of all placental mammals, the CpG island was not methylated and the *ZFX* gene escaped X inactivation. Then, *ZFX* became subject to CpG island methylation and X inactivation during the evolution of myomorph rodents, after the divergence of the sciuromorph (squirrel) lineage. This single event readily accounts for the pattern of X inactivation of *ZFX* among all eutherian species examined, representing 11 Orders.

Southern blotting experiments had established that humans have a highly conserved Y homolog of *ZFX* and that mice do not (Page et al., 1987). More widely, which mammalian orders exhibit a highly conserved Y homolog, and which do not? High stringency Southern blotting was carried out with one probe derived from the CpG island and a second probe derived from the zinc-finger domain of the human *ZFY* gene. Using either probe, a highly conserved, male-specific homolog appeared in 10 of the 13 eutherian species tested (figure 2.6; see also Page et al., 1987). The only species which lacked highly conserved Y homologs were the myomorph rodents (figure 2.6).

We conclude that in most mammalian lineages the CpG island and coding sequences of *ZFY* have diverged little from those present in the ancestor of all placental mammals. These lineages are also those in which *ZFX* continued to escape X inactivation. Only in the myomorph lineage, subsequent to the divergence of

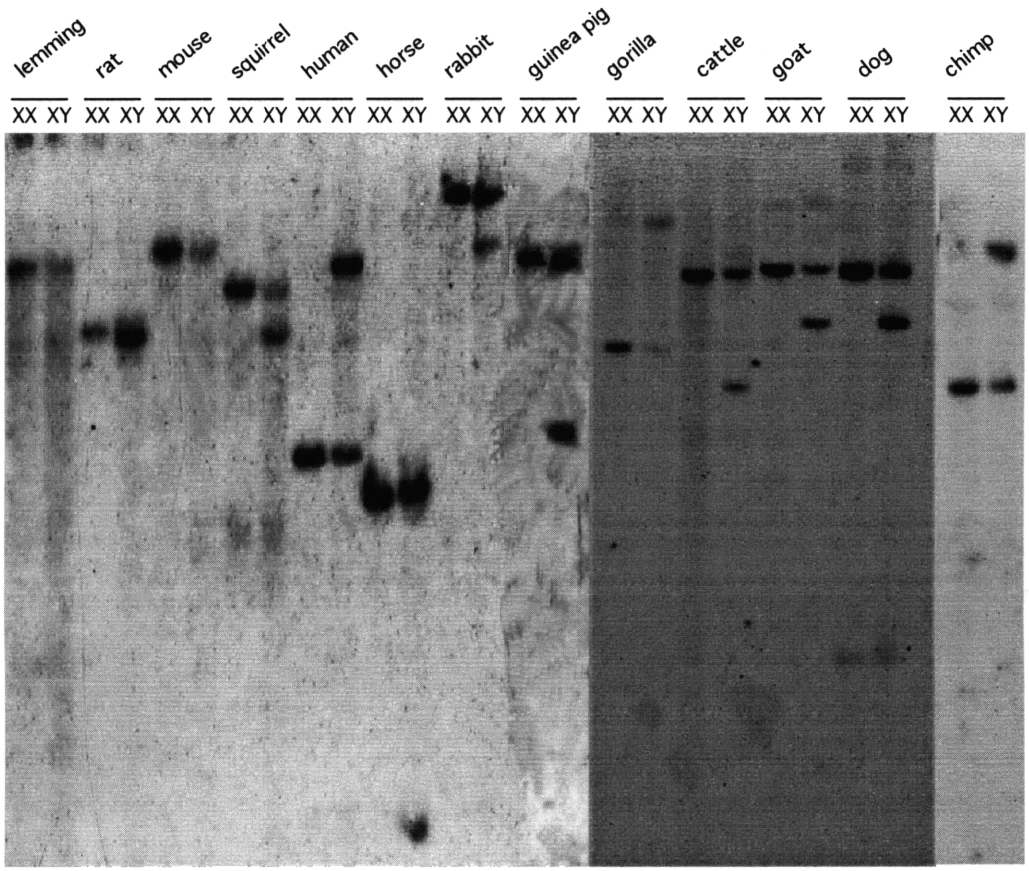


Figure 2.6 Closely related, Y-chromosomal homologs of *ZFX* detected in non-myomorph mammals by Southern blotting. Hybridization was with a 395 bp probe derived from the human ZFY CpG island.

sciuromorphs, did Y-chromosomal sequences evolve rapidly—perhaps reflecting a relaxation of functional constraints on *ZFY* genes in myomorphs.

Discussion

This study of *ZFX* and *ZFY* in the course of evolution provides a glimpse into the history of the sex chromosomes and of dosage compensation in placental mammals. In the ancestor of all eutherians, the X and Y chromosomes carried similar but nonidentical *ZFX* and *ZFY* genes, which had already diverged, but only slightly from a single common founder gene that may have moved to the sex chromosomes from an autosomal location (where homologs persist in marsupials and monotremes) (Sinclair et al., 1988; Spencer et al., 1991). In this eutherian ancestor, *ZFX* escaped X inactivation. This ancestral state of affairs has been maintained in most extant eutherian lineages. During myomorph evolution, however, significant changes befell the status of both *ZFX* and *ZFY*. About 40 to 70 million years ago—after the divergence of sciuromorphs but before the myomorph radiation (O'h Uigin and Li, 1992)—*ZFX* became subject to X inactivation and *ZFY* began to evolve at a much more rapid rate (figure 2.5).

For *ZFX* and *ZFY*, different groups of mammals equalize gene expression in males and females by using two fundamentally different strategies. The primitive solution, as exemplified by humans, is the expression of two gene copies per cell: in XX cells, two *ZFX* alleles, and in XY cells, one *ZFX* allele and one *ZFY* (highly similar in nucleotide sequence and pattern of expression). The more recent, derived solution, exemplified by mice, is the expression of one gene copy per cell: one *ZFX* allele in most XX or XY cells (the *ZFY* genes only expressed in testes).

The incremental changes in *ZFX* and *ZFY* that occurred during myomorph evolution illustrate broad principles at work in the evolution of sex chromosomes. The X and the Y are thought to have once been identical and that genes on the Y were lost while homologs on the X became subject to dosage compensation. This study

documents an instance of the degeneration of a Y chromosomal partner of an X-linked gene in a discrete, molecularly defined context. In the case of *ZFY* evolutionary change—observed in the myomorphs—indeed has been in the direction of diminished or restricted Y function. Degeneration of the Y thus appears to be ongoing, even among mammals, which are usually considered to display extreme sex chromosome heteromorphism.

Moreover, as shown, this incremental degeneration of the myomorph Y has been accompanied by dosage compensation—X inactivation—of the homologous gene on the X chromosome. In attempting to reconstruct the evolution of sex chromosomes, researchers have traditionally selected invertebrates or non-mammalian vertebrates for examples of transitional states. The findings on *ZFX* and *ZFY* evolution illustrate that transitional states in both X dosage compensation and Y degeneration exist at the level of individual genes among extant mammals. Escape of *ZFX* from X inactivation is not a novel invention in humans, but is instead a retained, primitive characteristic.

References

Adler, D. A., Bressler, S. L., Chapman, V. M., Page, D. C., and Disteche, C. M. (1991). Inactivation of the *Zfx* gene on the mouse X chromosome. *Proc Natl Acad Sci USA 88*, 4592-4595.

Cantrell, M. A., Bicknell, J. N., Pagon, R. A., Page, D.C., Walker, D. C., Saal, H. M., Zinn, A. R., Disteche, C. M. (1989). Molecular analysis of 46, XY females and regional assignment of a new Y-chromosome-specific probe. *Hum Genetics 83*, 88-92.

Feinberg, A. P., and Vogelstein, B. (1984). Addendum: A technique for radiolabeling DNA restriction endonuclease fragments to a high specific activity. *Anal Biochem 137*, 266-267.

Fialkow, P. J. (1970). X-chromosome inactivation and the Xg locus. *Am J Hum Genet 22*, 460-463.

Goodfellow, P., Pym, B., Mohandas, T., and Shapiro, L. J. (1984). The cell surface antigen locus, MIC2X, escapes X-inactivation. *Am J Hum Genet 36*, 777-782.

Luoh, S. W., Bain, P. A., Polakiewicz, R. D., Goodheart, M. L., Gardner, H., Jaenisch, R., and Page, D. C. (1997). Zfx mutation results in small animal size and reduced germ cell number in male and female mice. *Development 124*, 2275-2284.

Luoh, S. W., Jegalian, K., Lee, A., Chen, E. Y., Ridley, A., and Page, D. C. (1995). CpG islands in human *ZFX* and *ZFY* and mouse *Zfx* genes: sequence similarities and methylation differences. *Genomics 29*, 353-363.

Mardon, G., Luoh, S.-W., Simpson, E. M., Gill, G., Brown, L. G., and Page, D. C. (1990). Mouse Zfx protein is similar to Zfy-2: each contains an acidic activating domain and 13 zinc fingers. *Mol Cell Biol 10*, 681-688.

Mardon, G., and Page, D. C. (1989). The sex-determining region of the mouse Y chromosome encodes a protein with a highly acidic domain and 13 zinc fingers. *Cell 56*, 765-770.

Nagamine, C. M., Chan, K. M., Kozak, C. A., and Lau, Y. F. (1989). Chromosome mapping and expression of a putative testis-determining gene in mouse. *Science 243*, 80-83.

O'h Uigin, C., and Li, W. H. (1992). The molecular clock ticks regularly in muroid rodents and hamsters. *J Mol Evolution 35*, 377-384.

Page, D. C., Mosher, R., Simpson, E. M., Fisher, E. M., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A., and Brown, L. G. (1987). The sex-determining region of the human Y chromosome encodes a finger protein. *Cell 51*, 1091-1104.

Schneider-Gädicke, A., Beer-Romero, P., Brown, L. G., Mardon, G., Luoh, S.-W., and Page, D. C. (1989). Putative transcription activator with alternative isoforms encoded by human *ZFX* gene. *Nature 342*, 708-711.

Schneider-Gädicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R., and Page, D. C. (1989). *ZFX* has a gene structure similar to *ZFY*, the putative human sex determinant, and escapes X inactivation. *Cell 57*, 1247-1258.

Shapiro, L. J., Mohandas, T., Weiss, R., and Romeo, G. (1979). Non-inactivation of an x-chromosome locus in man. *Science 204*, 1224-1226.

Shimmin, L. C., Chang, B. H., and Li, W. H. (1993). Male-driven evolution of DNA sequences. *Nature 362*, 745-747.

Sinclair, A. H., Foster, J. W., Spencer, J. A., Page, D. C., Palmer, M., Goodfellow, P. N., and Graves, J. A. (1988). Sequences homologous to ZFY, a candidate human sex-determining gene, are autosomal in marsupials. *Nature 336*, 780-783.

Singer-Sam, J., and Riggs, A. D. (1993). X chromosome inactivation and DNA methylation, 358-383, in *DNA Methylation: Molecular Biology and Biological Significance*, Jost, J. P., and Saluz, H. P., eds. Birkhauser Verlag; Basel, Switzerland.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Evolution 98*, 503-517.

Spencer, J. A., Sinclair, A. H., Watson, J. M., and Graves, J. A. (1991). Genes on the short arm of the human X chromosome are not shared with the marsupial X. *Genomics 11*, 339-345.

Zambrowicz, B. P., Findley, S. D., Simpson, E. M., Page, D. C., and Palmiter, R. D. (1994). Characterization of the murine Zfy1 and Zfy2 promoters. *Genomics 24*, 406-408.

Zambrowicz, B. P., Zimmermann, J. W., Harendza, C. J., Simpson, E. M., Page, D. C., Brinster, R. L., and Palmiter, R. D. (1994). Expression of a mouse Zfy-1/lacZ transgene in the somatic cells of the embryonic gonad and germ cells of the adult testis. *Development 120*, 1549-1559.

Chapter 3. *RPS4*: On the Verge of Acquiring X inactivation or a Case of Reversion?

Background

RPS4 genes encode one of the approximately 33 proteins that constitute the small subunit of the ribosome, a massive structure that in eukaryotes contains four species of RNA and about 80 separate proteins (Wool, 1979). The RPS4 protein is located at the interface of the 40S and 60S subunits (Nygård and Nika, 1982; Uchiumi et al., 1986), associates with the initiation factor eIF-3 and may form part of the domain in the ribosome that binds mRNA (Westermann and Nygård, 1983). The genes that encode the ribosomal proteins are distributed throughout the eukaryotic genome. Each ribosomal protein gene in mammals has given rise to many processed pseudogenes—perhaps about 20 per gene. The pseudogenes are not expressed, however, so as a rule one gene encodes each ribosomal protein (Davies et al., 1989). In humans, where two *RPS4* genes exist, one on the X and one on the Y, both ubiquitously expressed, a male apparently has two different kinds of ribosomes, those which include the X isoform and those that include the Y isoform. The X and Y forms of human RPS4 are 93% conserved at the amino acid level (19 amino acid substitutions distinguish the 263 aa proteins from each other) (Fisher et al., 1990), and the Y homolog has been shown to be functionally interchangeable with the X gene (Watanabe et al., 1993).

RPS4X escapes X inactivation in humans, as shown by comparison of transcription levels in lymphoblastoid cell lines containing one, two, or four X chromosomes. The cell lines were derived from normal XY and XX individuals and from individuals with X chromosome aneuploidies, like XXXX and XXXXY. That human *RPS4X* is transcribed from the inactive X was confirmed by studying expression in human-rodent hybrid cell lines that segregate the inactive and active X chromosomes (Fisher et al., 1990).

A Y homolog of *RPS4* was reported in humans at the same time as the X homolog. Soon thereafter a mouse homolog of *RPS4* was found on its X chromosome, but thorough searches, by screening cDNA libraries, revealed no additional, no Y-linked genes in mice (Zinn et al., 1994). Thus, the case of *RPS4* parallels the case of *ZFX*. The gene is X inactivated in mice but not in humans. A conserved Y homolog exists in humans but not in mice. In the case of *RPS4* the Y homolog situation in mice is simpler than with *ZFX*: no expressed *RPS4Y* homolog exists at all while in the case of *ZFX*, as described before, Y homologs do exist in the mouse, but these do not appear to be functionally interchangeable with the X-linked gene. Not only are the 2 mouse *Zfy* genes relatively diverged in sequence, they are only expressed in testes in contrast to the ubiquitously expressed *ZFX* genes and human *ZFY*. The case of *RPS4* is more straightforward. Two copies of the *RPS4* gene are expressed in each human cell, whether male (where the second copy comes from the Y) or female (where the second copy comes from the inactivated X). However, one copy is expressed in the somatic cells of mice, whether male or female (Zinn et al., 1991). This state of affairs basically parallels that observed for *ZFX*.

While all the other known human genes that are not X inactivated and that have Y homologs map to the short arm of the X, *RPS4X* maps to the long arm of the X chromosome (Fisher et al., 1990). Indeed, *RPS4X* maps near the X inactivation center and the *XIST* gene, the locus from which the X inactivation signal that inactivates the chromosome emanates. Like *ZFX/ZFY*, *RPS4X/RPS4Y* appear to have diverged from a single common ancestral gene prior to the radiation of placental mammals. As between *ZFX* and *ZFY*, the homology between *RPS4X* and *RPS4Y* is basically limited to coding exons, where divergence would be selectively disadvantageous if the proteins retain their ancestral functions; meanwhile, introns and flanking sequences have diverged and can rarely be aligned. A comparison of silent nucleotide divergence rates, corrected for back mutations (Rice, 1987) revealed a four-fold greater divergence between the coding

regions of human *RPS4X* and *RPS4Y* than between the zinc finger domains of human *ZFX* and *ZFY*, suggesting that the *RPS4X/RPS4Y* gene pair has been diverging for longer than the *ZFX/Y* pair. Recall that the *ZFX* homolog is autosomal in monotremes and marsupials and thus seemingly has been acquired relatively recently by the mammalian sex chromosomes, after the divergence of marsupials from the placental mammalian lineage.

Since the X inactivation and Y homolog states observed for *RPS4* in humans and mice appear so closely to parallel what is observed with *ZFX* in humans and mice, it seemed reasonable to suppose that the two genes have followed similar evolutionarily trajectories through eutherian history. *ZFX* apparently escaped X inactivation and had a Y homolog in the eutherian ancestor and then acquired X inactivation and lost the Y homolog only in the rodent lineage. Did *RPS4* go through similar events? Perhaps some discrete change led to greater thoroughness of X inactivation in myomorphs and swept down *RPS4* along with *ZFX*.

Methods

Cloning and sequencing RPS4 intron 1 sequences. *RPS4* intron 1 PCR products were obtained using conditions described for PCR in Chapter 2 and the following primers: CCTA(G/A)CGCAGCCATGGTAAG, which overlaps with the first exon, and TCTTGGGACCACGAGCCT, which lies near the 5' junction of exon 2. To clone the intron from dog, the upstream primer used was GCCTCGCGCAGCCATGGT. Products were cloned into pCR™II vector using the TA Cloning Kit (Invitrogen). Plasmids were prepared for use as sequencing template using Wizard™ Maxipreps DNA Purification System (Promega). Sequencing was performed on an ABI 373 machine by dye-terminator method as prescribed by the manufacturer (Applied Biosystems).

Methylation assay. Conditions were as described for *ZFX* in Chapter 2. The primers used for each species are shown in figure 3.1.

cDNA selection. A biotinylated human *RPS4Y* coding sequence probe (selector) was hybridized at 55°C overnight against cDNA from male cow, dog, rabbit, or rat. Streptavidin coated beads were used to isolate what had bound to the selector. Streptavidin-biotin binding was allowed to occur in 10 mM Tris pH 7.5, 1 mM EDTA, 1 M NaCl. The beads were washed 2 x 15 min at room temperature and then 3 x 10 min at 55° in 1x SSC, 0.1% SDS. Selection products were eluted with 50 mM NaOH. The selection eluate was PCR amplified. A second round of selection was carried out. Products were cloned and sequenced as with *RPS4* intron 1 products above.

cDNA library screening. A dog cDNA library made from the liver of an adult male was obtained from Clontech. The cloning vector was λgt10. The library had been amplified once in C600 *Hfl.* Some 10^6 plaques were screened at low stringency with a probe derived from human *RPS4Y* coding sequence: the probe was hybridized to nylon membranes, lifted from library plates, in Church solution at 58°C overnight. The filters were washed 3 x 20 minutes in 1X SSC, 0.1% SDS at 50°C before being exposed to autoradiographic film at -80°C with an intensifying screen for a day. A secondary set of platings was performed to isolate positive clones. Once dog *RPS4X* was obtained, it was used as a probe for high stringency rescreening of the same library: this time hybridization was performed at 65°C and washes were at 65° in 0.1X SSC, 0.1% SDS.

An opossum (*Monodelphis domestica*) cDNA library made from the spleens of 5 males, cloned into the λZAPII vector, and amplified once was obtained from Stratagene. It was screened like the dog library except the high stringency screen was performed using an opossum *RPS4X* cDNA.

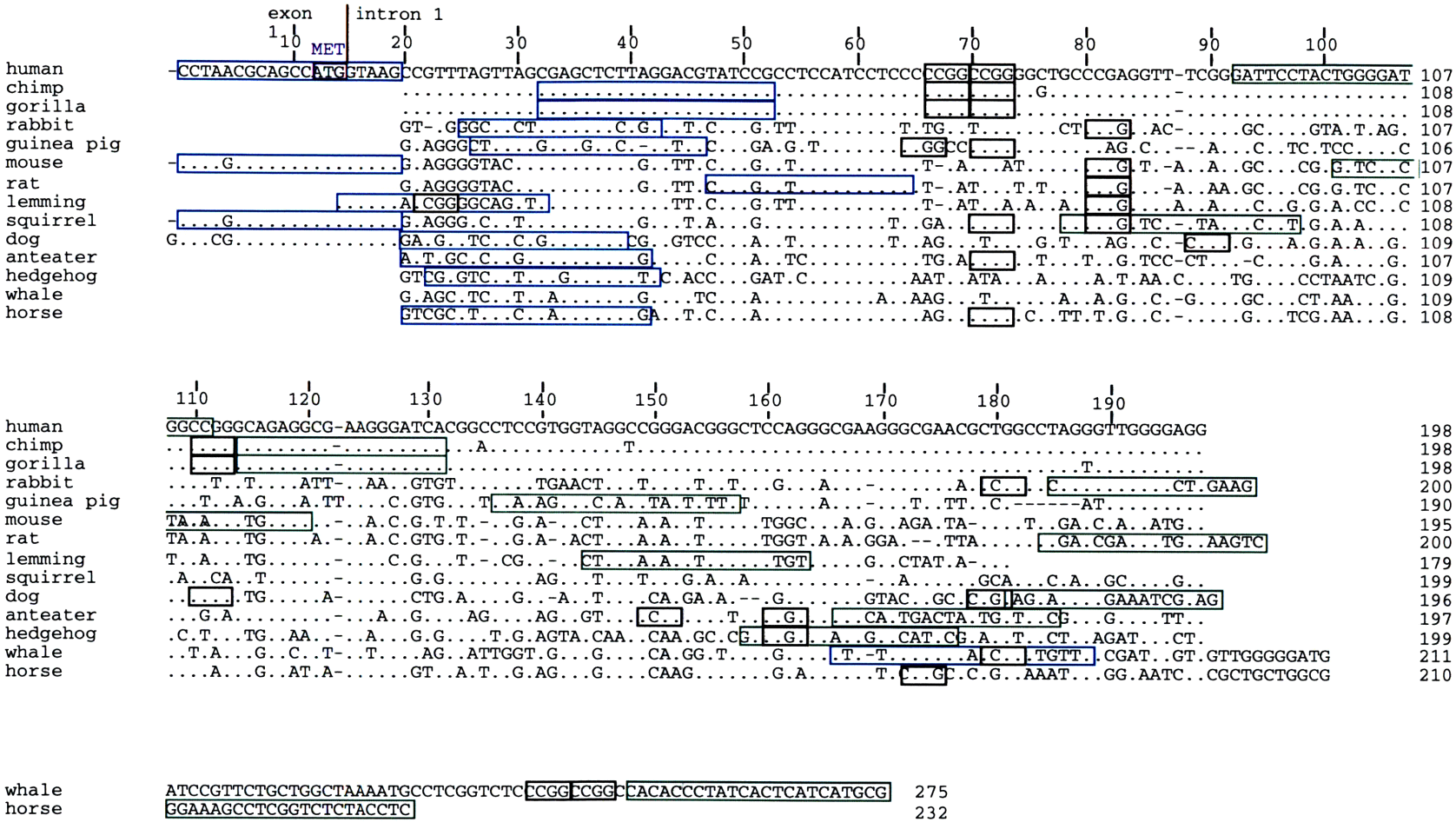


Figure 3.1 Fourteen species alignment of *RPS4X* showing the 5' portions of first introns. The start codon, the exon 1 to intron 1 boundary are noted. HpaII/MspI (CCGG) sites tested are shown in black boxes; PCR primers used in methylation assays are also shown: left primers are outlined in pale blue and right primers in green.

Results

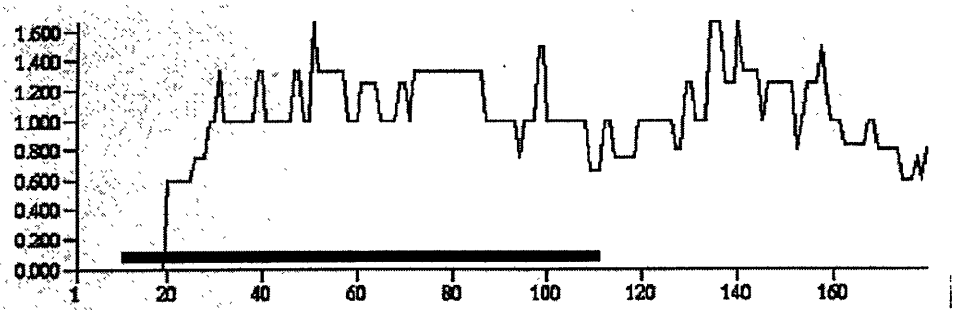
All known housekeeping genes have 5' CpG islands, and so does the ribosomal protein gene, *RPS4X* (figure 3.2). Its CpG island is a rather typical one, not nearly as conserved as the extraordinary *ZFX* CpG island. A single methylation assay could not be applied to a broad range of species as with *ZFX*, so to use CpG island methylation as a marker for X inactivation, I needed to develop unique ways to test methylation for individual species. Ultimately, many specific tests were used to survey a broad sweep of eutherians.

To develop species-specific (or at best order-specific) PCR-based methylation assays, I cloned and sequenced CpG island regions from different species mostly by one technique. The *RPS4X* CpG island overlaps into the gene's first intron, which is about 1 kilobase long in every tested mammal. Ribsomal protein genes are well conserved in coding sequence, as one would expect for such a basic component of the cellular machinery. I sought to clone the first introns of *RPS4X* genes, in order to obtain some CpG island sequences, using PCR with primers overlapping conserved portions of the bordering exons, However, the first exon of most ribosomal protein genes is tiny, including just three coding nucleotides and little 5' untranslated sequence. Fortunately, a short stretch of sequence overlapping the first exon is conserved enough to permit the design of primers that, when paired with primers from exon 2, ultimately succeeded in amplifying first introns of *RPS4X* homologs from 9 of 11 eutherian orders subjected to this method. In the end, only the corresponding region from elephant remained uncloned. To obtain dog intron 1 sequence, a cDNA library was first screened to isolate dog's *RPS4X* transcript, and this transcript was then used to design refined primers specifically for the cloning of the *RPS4X* first intron in dog. A 14 species alignment that starts from the first exon of *RPS4X*, extends into the 5' portion of the first intron, and shows how methylation was tested in each species (including the relevant restriction enzyme digestion sites and the PCR primers employed) is shown in figure 3.1.

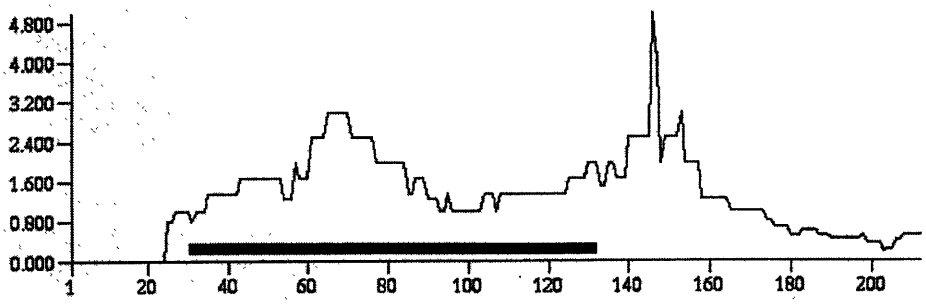
Figure 3.2 CpG to GpC dinucleotide ratios in the 5' regions, where methylation was assayed, of various species' *RPS4* first introns. In mammalian genomic DNA the CpG to GpC ratio tends to be near 0.2, but in CpG islands the ratio approaches 1. Ratios were calculated over 40 to 70 bp windows.

The graphs correspond to sequences shown in figure 3.1. Note that discrepancies are expected at the 5' and 3' ends of these kinds of graphs because dinucleotide counting occurs over windows. Irregularities are most pronounced when short sequences are analyzed. Regions over which PCR assays were designed are approximately indicated with bars over the *x* axes.

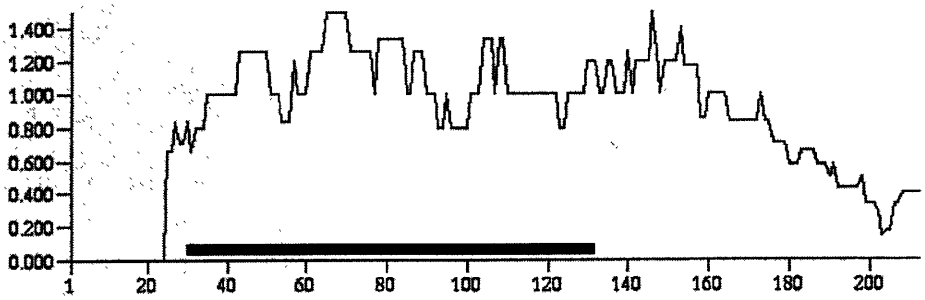
human



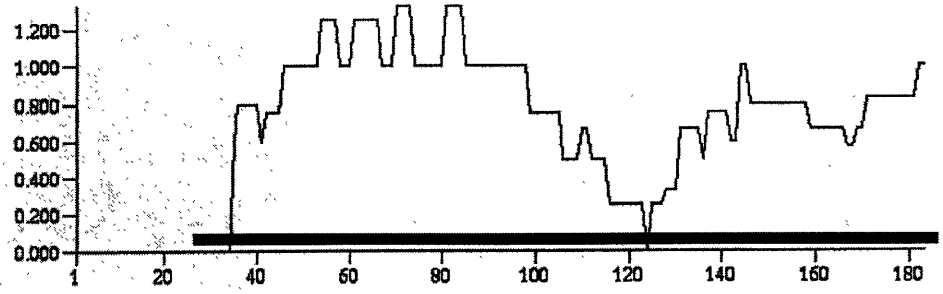
chimp



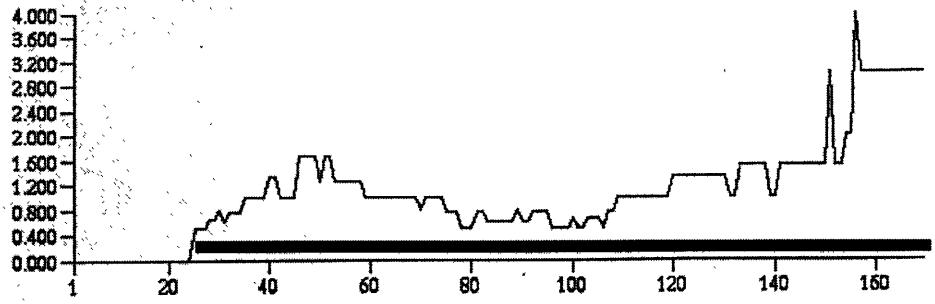
gorilla



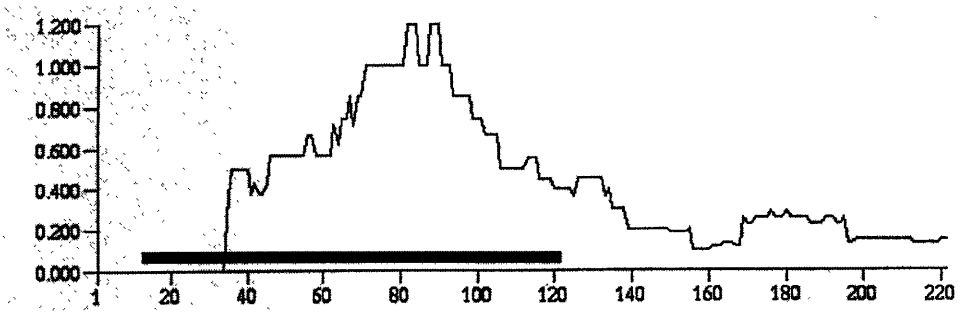
rabbit



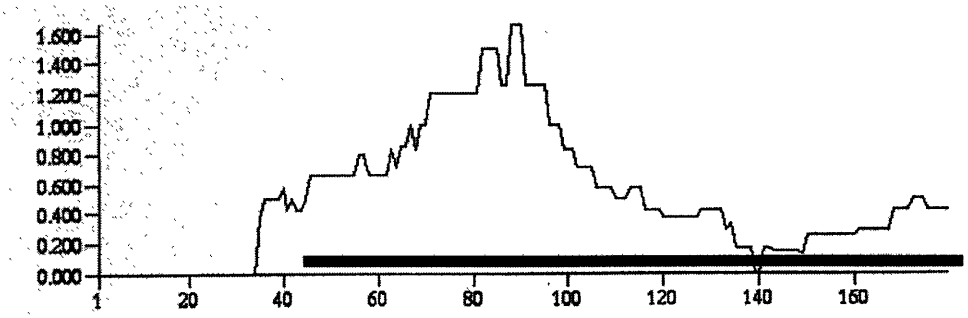
guinea pig



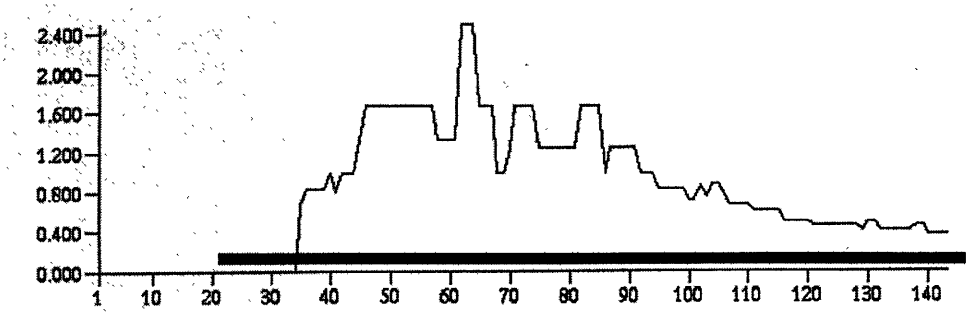
mouse



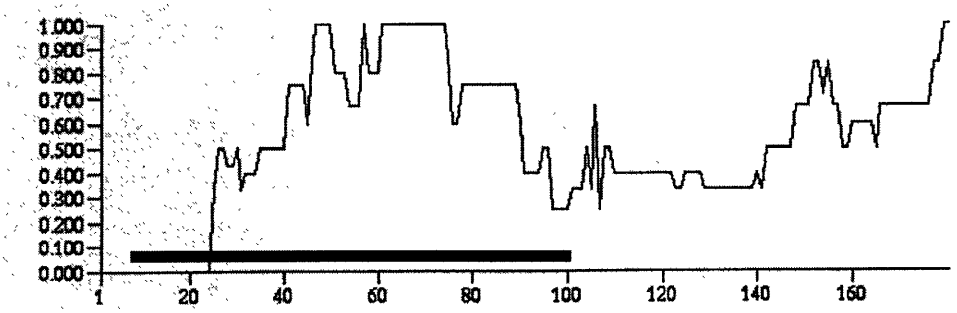
rat



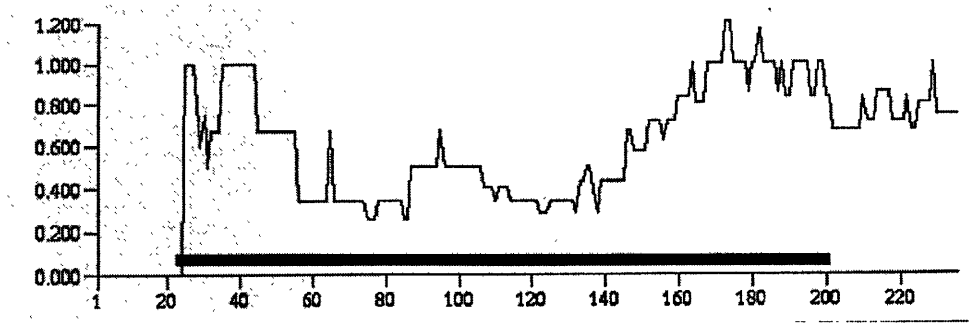
lemming



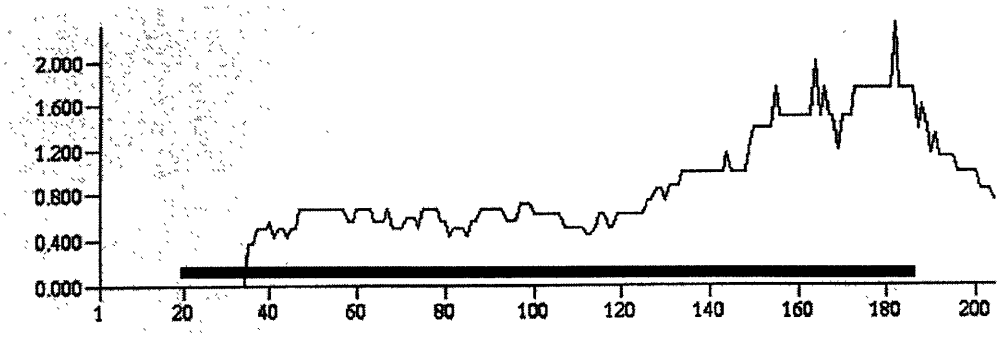
squirrel



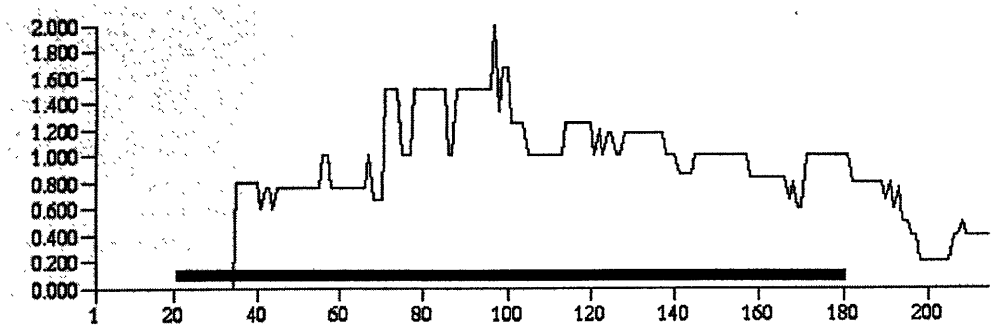
dog



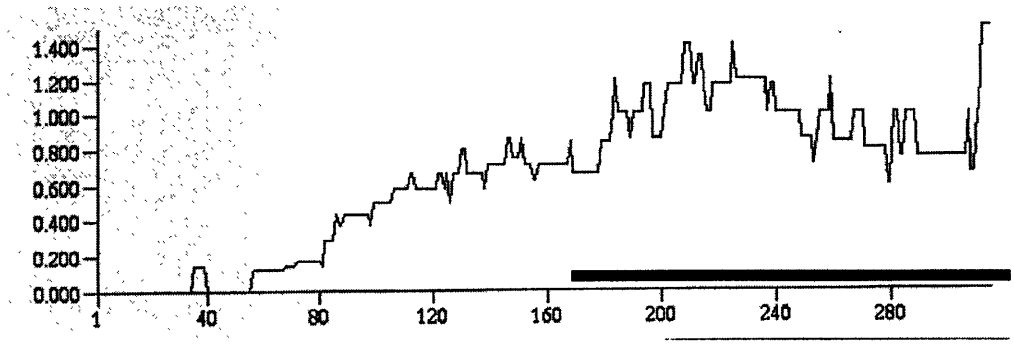
anteater



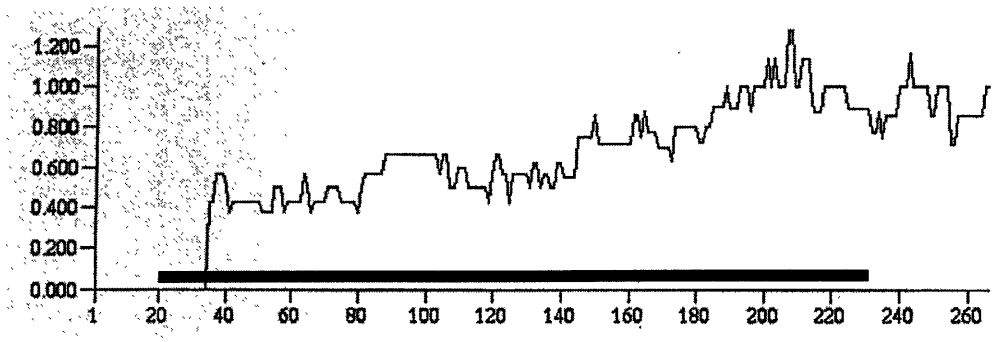
hedgehog



whale



horse



Instead of finding a pattern of X inactivation that mirrored the pattern observed with *ZFX*, I found something unexpected. As methylation results for more and more species trickled in, individually as species-unique assays were developed, it became increasingly clear that the primates, in which *RPS4X* escapes X inactivation, are outflanked by the many orders which show *RPS4X* methylation (figure 3.3). All tested rodents, including sciuromorphs, and representatives of the lagomorph, caviomorph, perissodactyl, carnivore, cetacean, edentate, and insectivore orders all showed methylation, implying X inactivation, of *RPS4X* in females. All tested males showed no methylation of the same region. (Artiodactyls were not tested because once their first introns were sequenced, no restriction sites amenable for methylation assays were found.) Only primates—all tested primates including apes, and Old World and New World monkeys—showed no 5' methylation, no X inactivation of *RPS4X*. (Despite repeated attentions, prosimians, the family of primates that first diverged from the lineage, remained perverse, and I could not develop an *RPS4X* methylation assay for them.) See figure 3.4 for all the results in the context of eutherian phylogeny.

To develop the story of *RPS4*, the evolution of Y homologs was also considered. But looking for Y homologs also was not as straightforward for *RPS4* as it was for *ZFX*, where Southern blot hybridization could reveal conserved male-specific sequences. The presence in the mammalian genome of often more than a dozen processed pseudogenes for each ribosomal protein gene obfuscates any attempt to screen for true *RPS4* genes by hybridization to genomic DNA.

A human Y homolog of *RPS4X* had first been detected as a 1 kb transcript on a Northern blot. The transcript was isolated from cDNA libraries, and the gene was shown to be present in males but not in females (Fisher et al., 1990). In mouse, cDNA library screening had failed to reveal any second, male-specific RPS4. I sought Y homologs in four more eutherians—rat, rabbit, cow, and dog—representing three additional orders, by cDNA selection. For each species, 24 cDNA clones selected by hybridization with

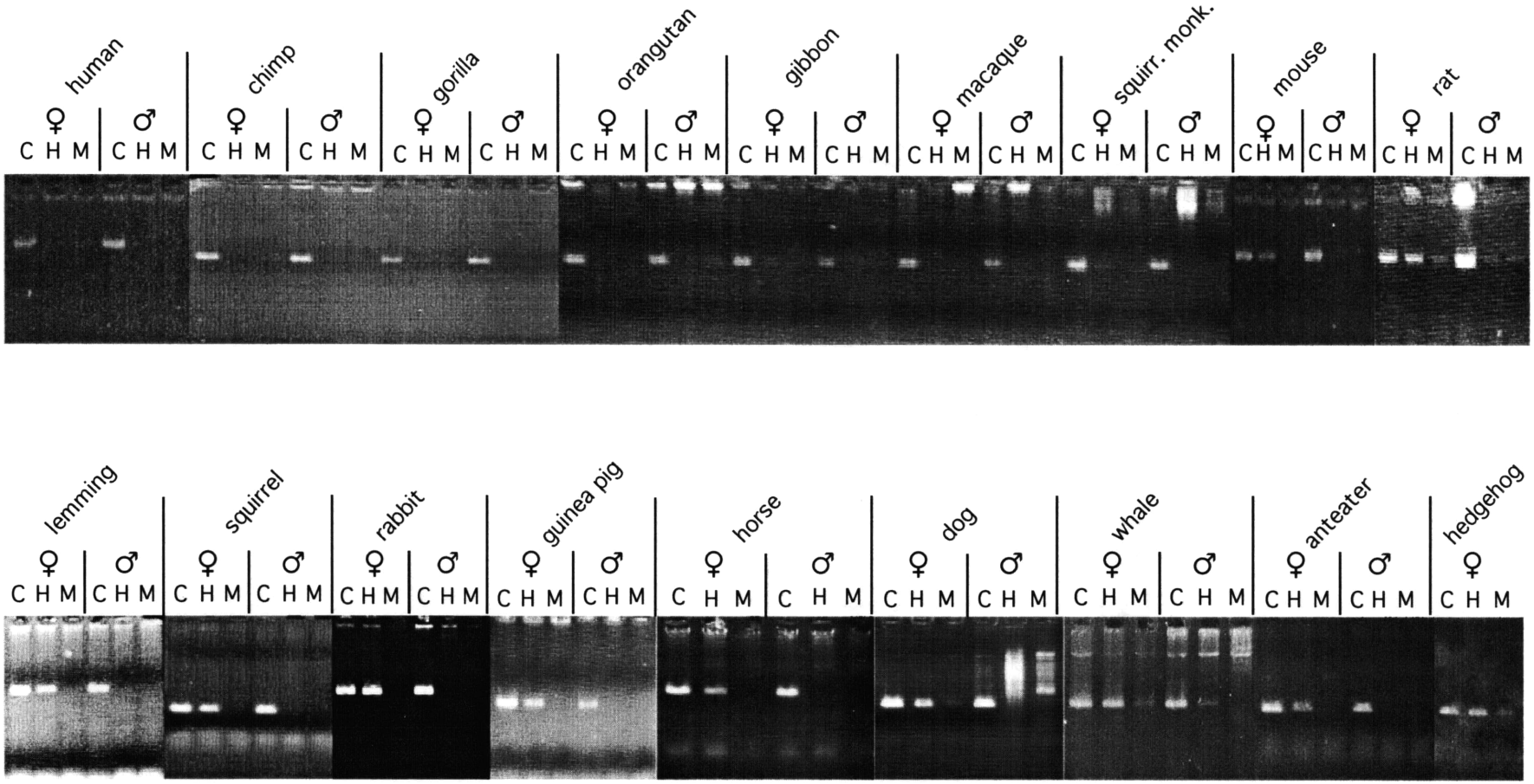


Figure 3.3 *RPS4X* methylation results. Female and male genomic DNAs digested with HpaII (H) or MspI (M) or HindIII (C, "control") were used as alternate templates in PCR assay diagrammed in figure 3.1. PCR products were visualized by ethidium bromide/UV staining after agarose gel electrophoresis.

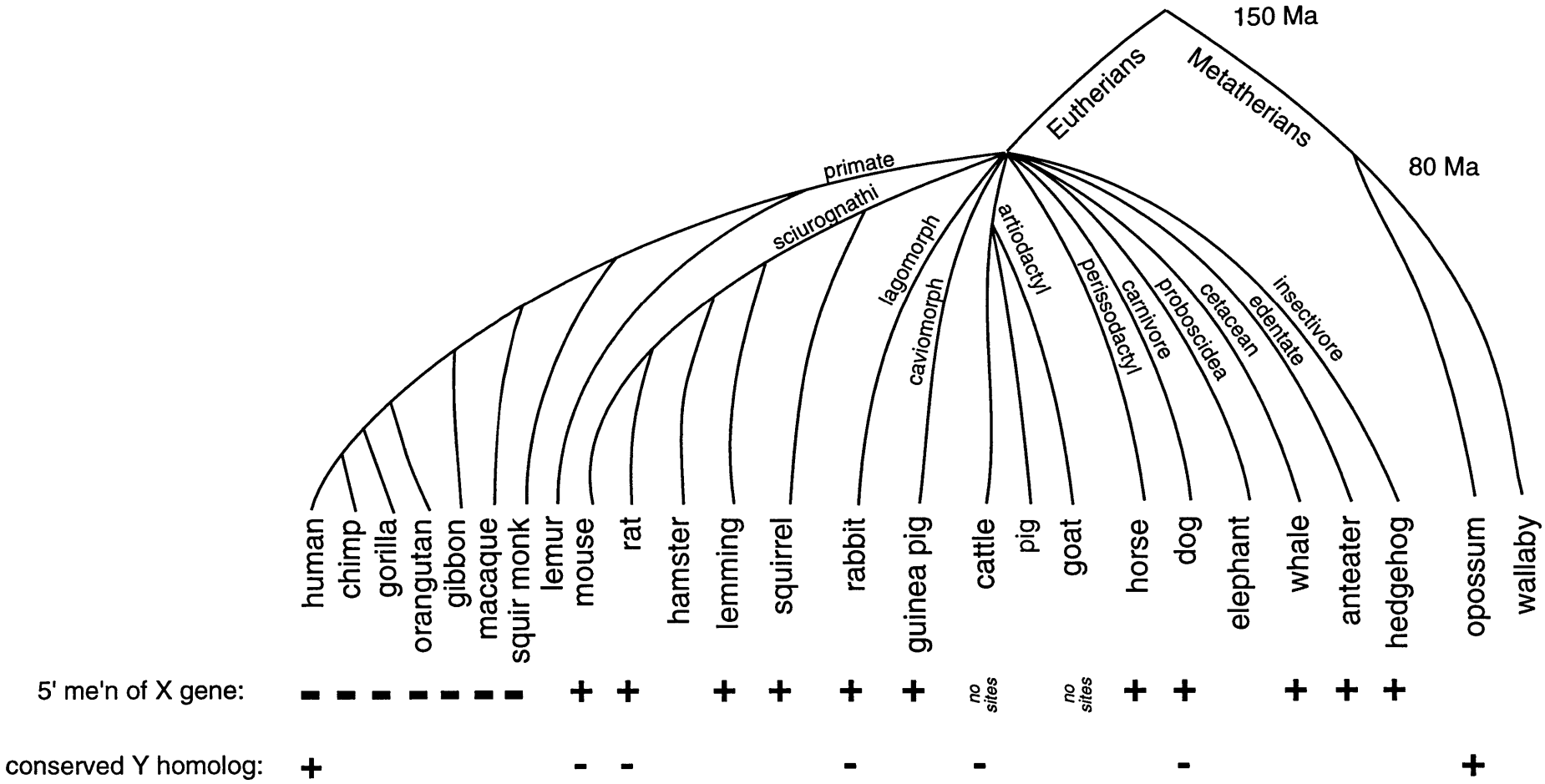


Figure 3.4 Results of *RPS4X* methylation and Y homolog studies in mammalian species, arranged phylogenetically. Specific methylation assays were designed for each family of eutherians. Y homologs were sought by cDNA screening (human, mouse, opossum), cDNA selection (rat, rabbit, cattle), or both (dog) using a human RPS4Y probe.

human *RPS4Y* were sequenced. In each of the four species, the results of the cDNA selection revealed clones that aligned into only one *RPS4*-homologous sequence class. Some clones within a species did show slight sequence variations, too rare to represent truly different genes and most likely artifacts introduced by the repeated rounds of PCR involved during cDNA selection. A Y homolog of *RPS4* was sought in dog by a second method as well: dog *RPS4X* was used to stringently rescreen the dog cDNA library from which dog *RPS4X* had been isolated by low stringency screening with a human *RPS4Y* probe; the arrays of clones observed upon high and low stringency screening were found to be identical, implying that no secondary, slightly diverged homologs of *RPS4X* exist in dog. (Greater explanation of this strategy follows, when opossum *RPS4Y* is discussed.) In all of the species where no Y homolog was detected, *RPS4* appears methylated, X inactivated in females, and all seemed to lack a Y homolog, or any second *RPS4* gene (figure 3.4). This state of affairs would lead to equal dosage between males and females, with one copy of *RPS4* expressed in each somatic cell.

What explanations could account for the widespread *RPS4* X inactivation and Y homolog absence among eutherians? *RPS4X* could have been X inactivated in the common ancestor of placental mammals and somehow reverted in the primates to escape from X inactivation. Or *RPS4* X inactivation could have been acquired independently many times in different eutherian lineages, in perhaps all but the primate order; perhaps conditions have been ripe for the acquisition of *RPS4* X inactivation and Y homolog loss. Finally, *RPS4* X inactivation could have been acquired just once in the eutherian lineage and excluded primates if primates diverged before the radiation of the other tested eutherian orders. These last two models really represent two extreme positions on a continuum, the earlier model framed more to suit the notion of bush-like eutherian radiation or late divergence of primates from the eutherian lineage. The earlier primates diverged from other eutherians, the fewer the instances of *RPS4* X inactivation acquisition (and Y homolog loss) need be invoked to account for the observed results.

The first model, that *RPS4X* was X inactivated in the eutherian ancestor and was reactivated in primates, is parsimonious—only one change in X inactivation status during eutherian evolution need be appealed to—but the model is problematic. A functionally interchangeable Y homolog exists in humans; if it existed in the eutherian ancestor as well, an ancestor would be postulated that had a conserved, widely expressed Y homolog yet inactivated the X homolog, a condition that has never been observed. If on the other hand, no Y homolog existed in the eutherian ancestor but was acquired on the Y, along with escape from X inactivation, during primate evolution, then human *RPS4X* and *RPS4Y* would not represent an ancient gene pair, long diverging on the sex chromosomes. To help distinguish between the model of ancestral activity from the model of ancestral inactivity of *RPS4*, an outgroup was sought. Marsupials are an outgroup to placental mammals. I set out to test whether a marsupial has two copies of *RPS4*, one on the X and one on the Y, or just one, an X or autosomal gene. If the former, that would be confirmation that human *RPS4X* and *RPS4Y* represent a pair of genes that have existed in distinct forms on the mammalian sex chromosomes for at least 150 million years, that is, since the divergence of marsupials and placental mammals. The existence of an *RPS4Y* in the eutherian ancestor (perhaps surviving in marsupials) would be predicted by the ancestral active model (the later models stated above) and not by the ancestral inactive model.

Some evidence already existed (stated in the Background section of this chapter) that human *RPS4X* and *RPS4Y* have been diverging for a long time, longer than *ZFX* and *ZFY* have been, and not just since the primates diverged from the other eutherians. Finding an *RPS4Y* homolog in a marsupial would not only show that *RPS4X* and *RPS4Y* have been evolving side by side but distinctly on the sex chromosomes for a long time, but the fact that an *RPS4Y survived* in opossum would support the view that *RPS4X* escaped X inactivation in ancestral mammals.

A male opossum, *Monodelphis domestica*, cDNA library was screened first at low stringency with a human *RPS4Y* coding sequence probe. A major species of *RPS4* cDNA was detected. The major species of opossum *RPS4* was next used to rescreen the opossum cDNA library, this time at high stringency. About 15% of the clones that had hybridized upon low-stringency screening failed to appear upon high-stringency screening with the opossum major *RPS4*. The clones that had disappeared on high-stringency screening were sequenced and proved to represent a second species of opossum *RPS4*. See figure 3.5 for a coding sequence alignment and figure 3.6 for an amino acid alignment of various *RPS4*s, including the two opossum species.

The major and minor species of *RPS4* found in opossum were mapped to the X and Y chromosomes respectively. An intron was cloned from the major species of *RPS4* and this fragment of genomic DNA was used as a Southern probe on a blot that included 3 male and 3 female opossums. The Southern hybridization showed a band that was twice as strong in females as in males, after normalizing for DNA amounts in the lanes (figure 3.7). Thus, the major *RPS4* represents an X-linked gene. Methylation is not associated with inactivation of marsupial X chromosomes (Kaslow and Migeon, 1987), so simple assessment of CpG island methylation of genomic DNA could not reveal the X inactivation status of *RPS4X* in marsupials.

The minor species of *RPS4* was mapped to the Y as follows to evade the problem of obfuscating pseudogenes. PCR assays were designed in the 5' and 3'-UTRs of the minor *RPS4* gene. The UTR regions of the two forms of opossum *RPS4* are so diverged as to be beyond alignment, so no PCR amplification could occur from the major *RPS4*. PCRs were performed on female and male opossum DNA templates that had been equalized in concentration. When a relatively low PCR annealing temperature was used, stronger amplification could be observed from the male template than from the female. As the PCR annealing temperature was ramped up, female amplification disappeared before male amplification did (figure 3.8). Product observed from the male presumably

Figure 3.5 Alignment of coding regions from human and opossum *RPS4X* genes, the chicken *RPS4* gene, and human and opossum *RPS4Y* genes.

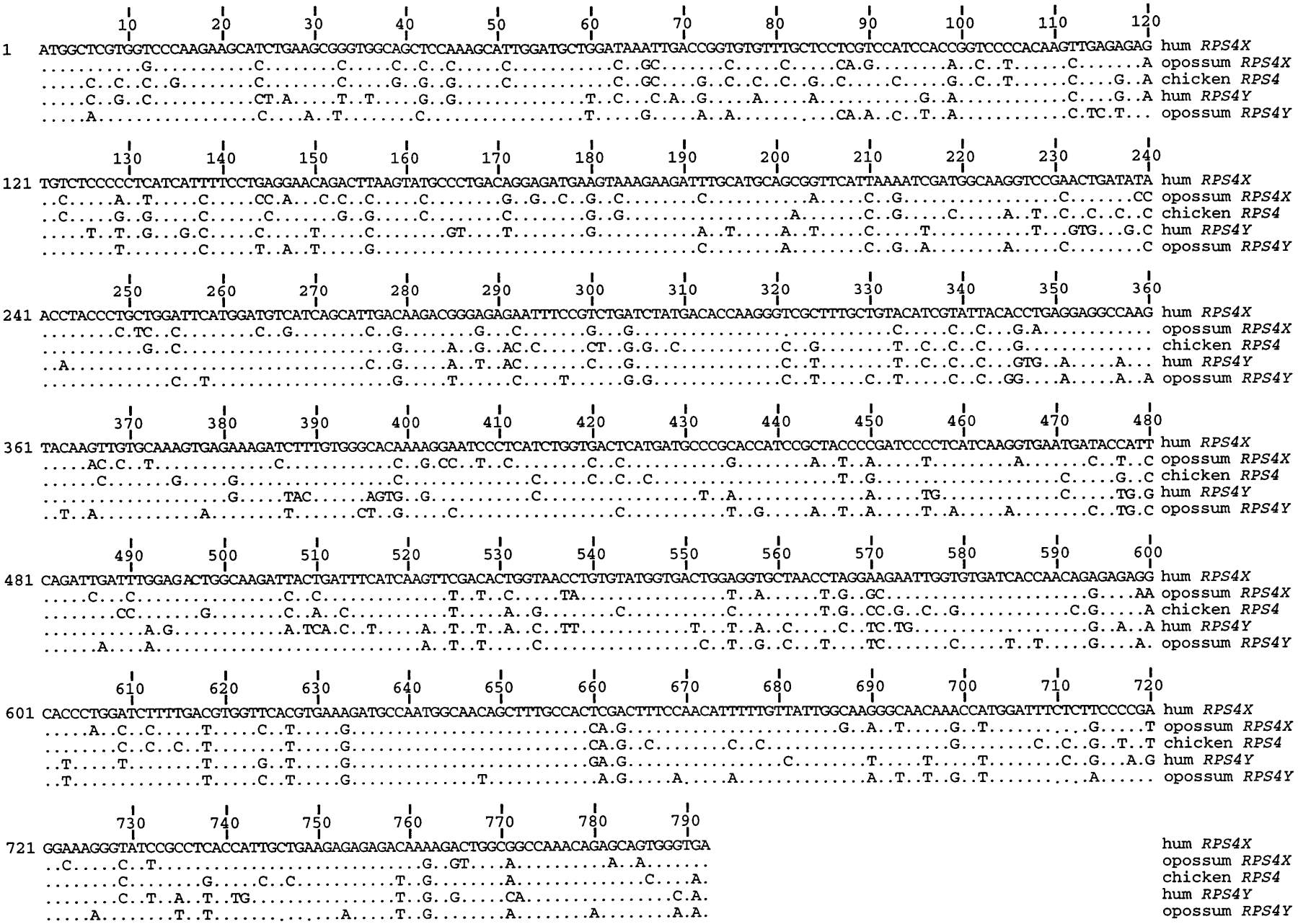
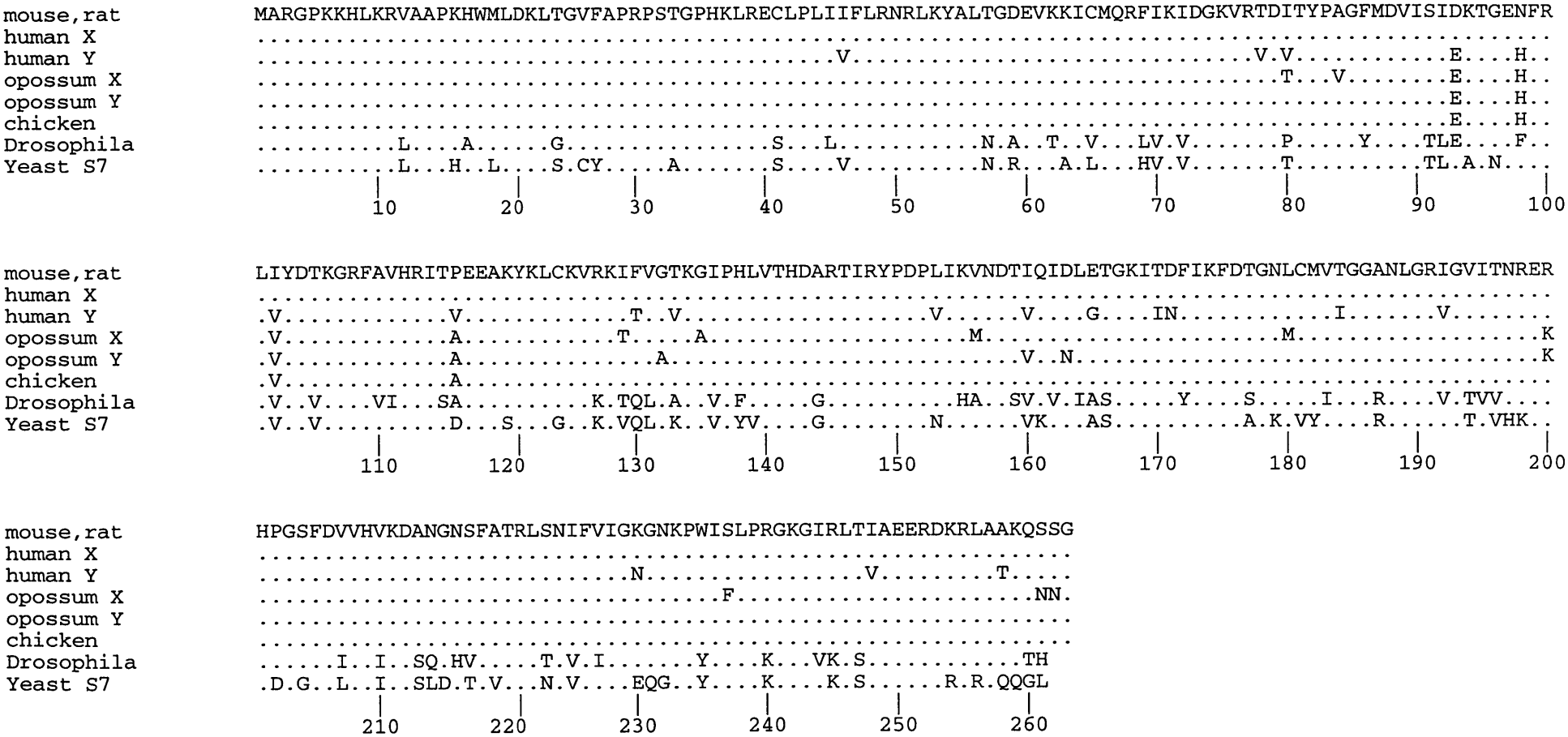

1 ATGGCTCGTGGTCCCAAGAAGCATCTGAAGCGGGTGGCAGCTCCAAAGCATTGGATGCTGGATAAATTGACCGGTGTGTTTGCTCCTCGTCCATCCACCGGTCCCCACAAGTTGAGAGAG hum RPS4X
.....G.....C.....C.....C..C..C.....C.....C..GC.....C.....C.....CA.G.....A..C..T.....C.....A opossum RPS4X
.....C..C..C..G.....C.....C.....G..G..G.....C.....C..GC....G..C..C..C..G..C.....C.....G..C..T.....C....G..A chicken RPS4
.....C..G..C.....CT.A.....T..T.....G..G.....T..C...C.A..G.....A.....A.....G..A.....C....G..A hum RPS4Y
.....A.....C....A..T.....C.....T.....G.....A..A.....CA.A..C..T..A.....C.TC.T... opossum RPS4Y
121 TGTCTCCCCCTCATCATTTTCCTGAGGAACAGACTTAAGTATGCCCTGACAGGAGATGAAGTAAAGAAGATTTGCATGCAGCGGTTCATTAAAATCGATGGCAAGGTCCGAACTGATATA hum RPS4X
..C.....A..T.....C.....CC.A...C.C..C.....C.....G..G..C..G..C.....C.....A.....C..G.....C.....CC opossum RPS4X
..C.....G..G.....C.....C.....G..G.....C.....C.....G..G.....A.....C..G.....C.....A..T..C..C..C..C chicken RPS4
.....T..T..G...G.C.....C.....T.....C.....GT....T.....G.....A..T.....A..T.....C.....T.....T...GTG...G.C hum RPS4Y
.....T.....C.....T..A..T.....G.....C.....A.....C..G..A.....A.....C.....C opossum RPS4Y
241 ACCTACCCTGCTGGATTCATGGATGTCATCAGCATTGACAAGACGGGAGAGAAATTTCCGTCTGATCTATGACACCAAGGGTCGCTTTGCTGTACATCGTATTACACCTGAGGAGGCCAAG hum RPS4X
.....C.TC..C.....C..G.....C..G.....G...C.....G...G.....C.....C..C...G.A..... opossum RPS4X
.....G..C.....G.....A..G..AC.C.....CT..G.G..C.....C..G.....T..C..C..C...G..... chicken RPS4
..A.....C..G.....A..T..AC.....C...G.....C..T.....T..C..C..C...GTG..A.....A... hum RPS4Y
.....C..T.....G.....T.....C....T.....G.G.....C..T.....C..T.....C..C..GG....A.....A..A opossum RPS4Y
361 TACAAGTTGTGCAAAGTGAGAAAGATCTTTGTGGGCACAAAAGGAATCCCTCATCTGGTGACTCATGATGCCCGCACCATCCGCTACCCCGATCCCCTCATCAAGGTGAATGATACCATT hum RPS4X
.....AC.C..T.....C.....C..G.CC..T..C.....C..C.....G.....A..T..A.....T.....A.....C..T..C opossum RPS4X
.....C.....G.....G.....C.....C.....C..C..C.....T..G.....C.....G..C chicken RPS4
.....G.....TAC.....AGTG..G.....C.....T..A.....A.....TG.....C.....TG.G hum RPS4Y
..T..A.....A.....T.....CT..G.....C.....C.....T..G.....A..T..A.....T..A.....A.....C..TG.C opossum RPS4Y
481 CAGATTGATTTGGAGACTGGCAAGATTACTGATTTCATCAAGTTCGACACTGGTAACCTGTGTATGGTGACTGGAGGTGCTAACCTAGGAAGAATTGGTGTGATCACCAACAGAGAGAGG hum RPS4X
.....C...C.....C..C.....T..T..C.....TA.....T..A.....T.G..GC.....G....AA opossum RPS4X
.....CC.....G.....C..A..C.....T.....A..G.....C.....C.....T.G..CC.G..C..G.....C.G.....A chicken RPS4
.....A.G.....A.TCA.C..T.....A..T..T..A..C..TT.....T...T..A..C.....C..TC.TG.....G..A..A hum RPS4Y
.....A....A.....A..T..T.....C.....C..T..G..C...T....TC.....C.....T..T.....G....A. opossum RPS4Y
601 CACCCTGGATCTTTTGACGTGGTTCACGTGAAAGATGCCAATGGCAACAGCTTTGCCACTCGACTTTCCAACATTTTTGTTATTGGCAAGGGCAACAAACCATGGATTTCTCTTCCCCGA hum RPS4X
.....A..C..C.....T.....C..T.....G.....CA.G.....G..A..T.....G..T.....G.....T opossum RPS4X
.....C..C..C..T.....T.....G.....CA.G..C.....C..C.....G.....C..C..G..T..T chicken RPS4
..T.....T.....T.....G..T.....G.....GA.G.....C.....T.....T.....T.....C..G...A.G hum RPS4Y
..T.....T.....C..T.....G.....T.....A.G.....A.....A.....A..T..T..G..T.....A..... opossum RPS4Y
721 GGAAAGGGTATCCGCCTCACCATTGCTGAAGAGAGAGACAAAAGACTGGCGGCCAAACAGAGCAGTGGGTGA hum RPS4X
..C.....C..T.....G..GT....A.....A..A..... opossum RPS4X
.....C.....G.....C..C.....T..G.....A.....C....A. chicken RPS4
.....C..T..A..T..TG.....T..G..G.....CA.....C.A. hum RPS4Y
.....A.....T..T.....A.....T..G.....A.....A.....A.A. opossum RPS4Y

Figure 3.6 Alignment of RPS4 protein sequences from rodents, human, chicken, opossum, Drosophila, and yeast S7.


mouse,rat MARGPKKHLKRVAAPKHWMLDKLTGVFAPRPSTGPHKLRECLPLIIFLRNRLKYALTGDEVKKICMQRFIKIDGKVRTDITYPAGFMDVISIDKTGENFR
human X
human Y
opossum X
opossum Y
chicken
Drosophila
Yeast S7
10 20 30 40 50 60 70 80 90 100
mouse,rat LIYDTKGRFAVHRITPEEAKYKLCKVRKIFVGTKGIPHLVTHDARTIRYPDPLIKVNDTIQIDLETGKITDFIKFDTGNLCMVTGGANLGRIGVITNRER
human X
human Y
opossum X
opossum Y
chicken
Drosophila
Yeast S7
110 120 130 140 150 160 170 180 190 200
mouse,rat HPGSFDVVHVKDANGNSFATRLSNIFVIGKGNKPWISLPRGKGIRLTIAEERDKRLAAKQSSG
human X
human Y
opossum X
opossum Y
chicken
Drosophila
Yeast S7
210 220 230 240 250 260

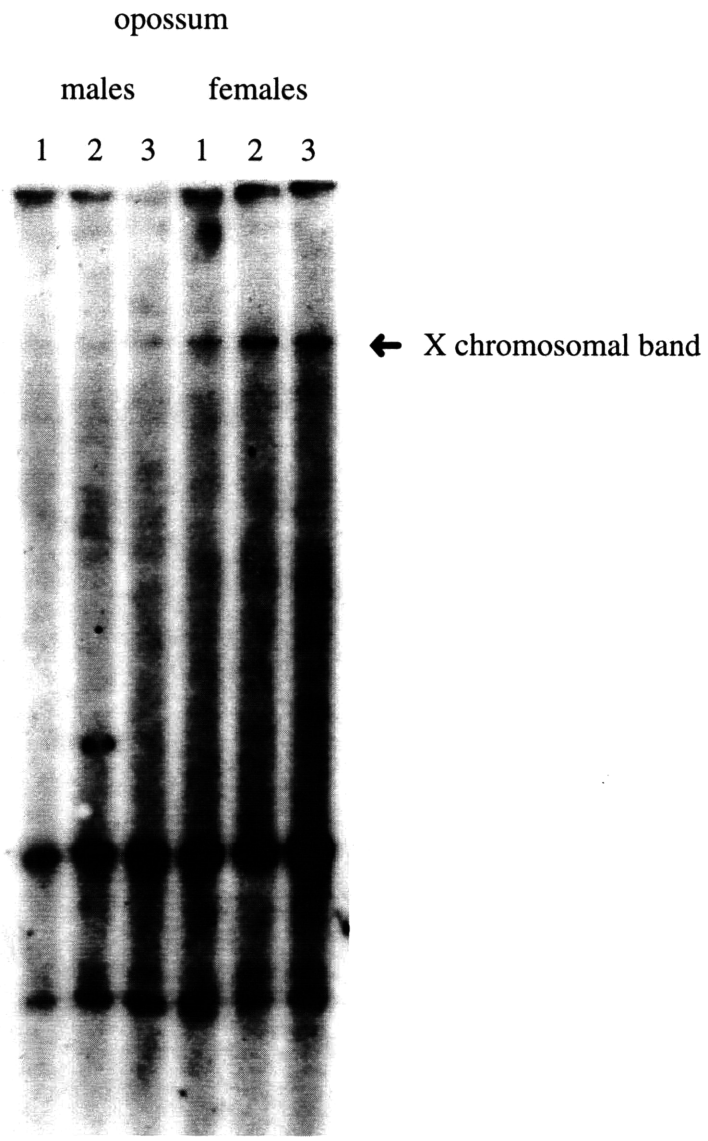


Figure 3.7 The opossum's major *RPS4* gene maps to the X chromosome. Southern hybridization against genomic DNAs was performed with the gene's sixth intron.

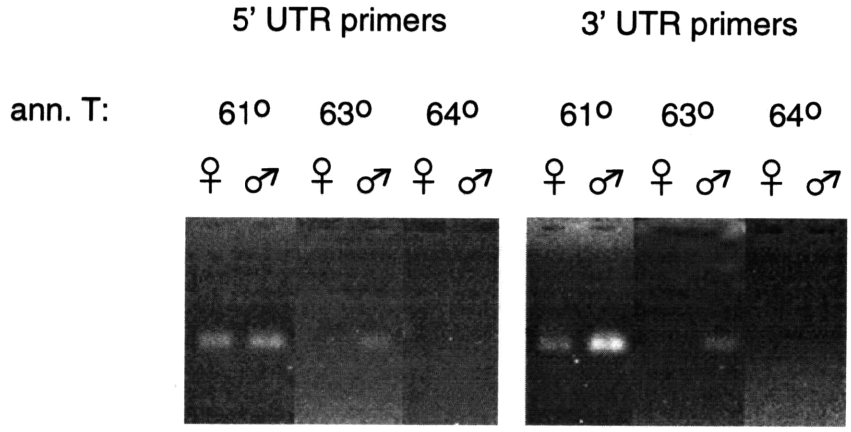


Figure 3.8 Mapping the opossum *RPS4* minor gene to the Y chromosome. PCR primers were picked from the gene's untranslated regions (UTRs). PCR products obtained when using female opossum DNA as PCR template arise from pseudogenes. As PCR annealing temperatures were raised, PCR products could be obtained from male DNA template at higher annealing temperatures than from female DNA template.

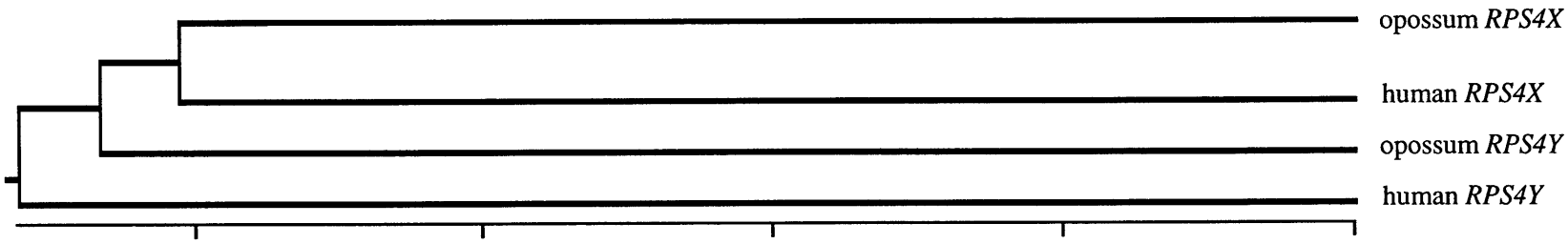


Figure 3.9 Phylogenetic tree showing the relationships of *RPS4* coding sequences at the nucleotide level (see figure 3.5). Note that the X-linked genes are more closely related to each other than to the Y-linked genes in the same species.

represents amplification, partly if not solely, from the real gene target while product observed from the female seems to arise from opossum *RPS4Y* pseudogenes, on the autosomes, that are just beginning to decay.

The presence of two distinct species of *RPS4*, one mapping to the X and one to the Y, in marsupials as well as in primates, implies that the X-linked and the Y-linked genes have existed in distinct forms for at least 150 million years, that is, prior to the time that marsupials diverged from eutherians and that *RPS4X* was likely to be active in ancestral mammals. The case of *RPS4X/RPS4Y* represents a third example of an X-Y gene pair common to both metatherian and eutherian sex chromosomes. Only *UBE1* (Wilcox et al., 1996) and *SMC* (Duffy and Graves, unpublished results) had been shown to map to the X and Y in both marsupials and humans.

Discussion

The human *RPS4X* and *RPS4Y* genes appear to be genuinely ancient residents of the sex chromosomes. They map to the X and Y in marsupials as well as in primates. Human *RPS4X* is positioned among markers on the X chromosome that are shared with the marsupial X (J. Graves, personal communication). In contrast, *ZFX* and *ZFY* and indeed most of the genes that escape X inactivation and have Y homologs, when tested, have been found to be autosomal in marsupials and monotremes (Graves, 1987; Graves and Watson, 1991). Only *SMCX*, *UBE1*, and *RPS4X*, among the genes that escape inactivation in humans, appear clearly to have been on the mammalian sex chromosomes all along. Most of the human Xp genes with Y homologs appear to have settled on the sex chromosomes after marsupials diverged from eutherians, though before the eutherian orders radiated. Phylogenetic relationships inferred from the alignment of *RPS4* genes (figure 3.9) are consistent with the notion of an early separation of X and Y homologs. The *RPS4X* and *RPS4Y* pair within opossum or within human are more diverged from each other than opossum and human *RPS4X* are from each other; the greater divergence

among the Y homologs is not unexpected in light of the more rapid evolution generally of Y chromosome genes (Shimmin et al., 1993).

So we know that *RPS4X* and *RPS4Y* have existed separately for at least 150 million years, that *RPS4* escapes X inactivation and retains a Y homolog in primates, but that in many other eutherian lineages *RPS4* is X inactivated and the Y homolog has been lost. How can these observations be explained? The eutherian ancestor seems clearly to have had an intact Y homolog of *RPS4* because the Y homolog is ancient and retained in at least some eutherians. If all examined eutherian orders but primates X inactivate *RPS4*, the most parsimonious explanation might be that *RPS4* was also X inactivated in the eutherian ancestor. In most eutherian lineages, then, *RPS4* X inactivation would have been maintained, leaving no selective pressure to retain the Y homolog. In primates *RPS4* X inactivation would have been lost under this scenario, keeping selective pressure to retain the Y homolog. Two difficulties beset this model: a eutherian ancestor would have X inactivated *RPS4X* while retaining *RPS4Y* and then X inactivation would have had to be lost after being acquired. X-Y homologous genes can clearly evolve independently, but this model calls for what seems evolutionarily unreasonable conduct.

Alternatively, *RPS4* could have escaped X inactivation in the common eutherian ancestor, and X inactivation could have been acquired in most eutherian orders independently, many times perhaps, either because the mechanistic barriers to acquisition of *RPS4* X inactivation were slight or, what seems more likely, selective pressures made *RPS4* ripe for the acquisition of X inactivation. Human *RPS4X* expression far outpaces *RPS4Y* expression, at a ratio of about 6 to 1 (Zinn et al., 1994). Interestingly, *RPS4Y* appears to be expressed in opossum at similarly reduced levels compared to the opossum *RPS4X*. Once the expression of a Y homolog is reduced, the pressure to guard against the inactivation of the X homolog may be reduced. Where a Y homolog is expressed at low levels, total expression from the X and the Y in a male will not be much higher than in a female with just one active X allele. This may have been the position of *RPS4* in the

eutherian ancestor, and most eutherian lineages began to X inactivate *RPS4*, eliminating any remaining selective pressure to retain the Y homolog.

A third explanation, really a variant of the second explanation, could account for the observed pattern of *RPS4* X inactivation and Y decay. If primates diverged from the eutherian lineage before the radiation of the other tested orders, *RPS4* X inactivation could have been acquired once, shortly after the divergence of the primates from the other eutherians. This explanation would account for what is observed without invoking either more than one event leading to acquisition of X inactivation or a reversion event in primates. Even if primates did not diverge first but did diverge relatively early from other eutherians, only a few independent changes in dosage strategy could explain the results. Rodents (sciurognathi *or* hystricomorphs), insectivores, and edentates have variously been called among the earliest to diverge from other eutherian orders. But the sequence of eutherian order divergence remains controversial, a question under continuing examination. No one order is universally considered the earliest to have diverged, so in the absence of strong evidence that places another order earlier in the sequence of divergence and in the absence of evidence that primates were not among the first eutherians to diverge, this third model to account for the observations of *RPS4* evolution cannot be discounted.

Most basically, the different patterns of X inactivation and Y decay observed with *RPS4* and with *ZFX/ZFY* show that these genes have independent histories and display unique transition states in the general evolution of the sex chromosomes toward increasing X inactivation and Y degeneration. The variablility in the specific evolution of these two sets of X-Y homologous genes shows that in the general tendency toward greater dosage compensation and Y chromosome degeneration, particular genes can evolve independently. Though *ZFX* and *RPS4* have evolved independently, viewing multispecies findings about either gene in phylogenetic context simplifies the findings and shows the validity of an evolution-minded examination of these genes' X inactivation

and Y homolog states. That events can be defined to explain the observations of X inactivation and Y homolog differences implies that the events leading to differences are infrequent. Otherwise, overlaying the results in an evolutionary context would not impose order onto the observations.

References

Davies, B., Feo, S., Heard, E., and Fried, M. (1989). A strategy to detect and isolate an intron-containing gene in the presence of multiple processed pseudogenes. *Proc Natl Acad Sci USA 86*, 6691-6695.

Fisher, E. M., Beer-Romero, P., Brown, L. G., Ridley, A., McNeil, J. A., Lawrence, J. B., Willard, H. F., Bieber, F. R., and Page, D. C. (1990). Homologous ribosomal protein genes on the human X and Y chromosomes: Escape from X inactivation and possible implications for Turner syndrome. *Cell 63*, 1205-1218.

Graves, J. A. (1987). The evolution of mammalian sex chromosomes and dosage compensation — clues from marsupials and monotremes. *Trends Genet 3*, 252-256.

Graves, J. A., and Watson, J. M. (1991). Mammalian sex chromosomes: evolution of organization and function. *Chromosoma 101*, 63-68.

Kaslow, D. C., and Migeon, B. R. (1987). DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation. *Proc Natl Acad Sci USA 84*, 6210-6214.

Nygård, O., and Nika, H. (1982). Identification by RNA-protein cross-linking of ribosomal proteins located at the interface between the small and the large subunits of mammalian ribosomes. *EMBO J 1*, 357-362.

Rice, K. A. (1987). Estimating nucleotide sequence divergence from polymorphism in nucleotide sequences and restriction maps. (Athens, Georgia: University of Georgia).

Shimmin, L. C., Chang, B. H., and Li, W. H. (1993). Male-driven evolution of DNA sequences. *Nature 362*, 745-747.

Uchiumi, T., Kikuchi, M., and Ogata, K. (1986). Cross-linking study on protein neighborhoods at the subunit interface of rat liver ribosomes with 2-iminothiolane. *J Biol Chem 261*, 9663-9667.

Watanabe, M., Zinn, A. R., Page, D. C., and Nishimoto, T. (1993). Functional equivalence of human X- and Y-encoded isoforms of ribosomal protein S4 consistent with a role in Turner syndrome. *Nat Genet 4*, 268-271.

Westermann, P., and Nygård, O. (1983). The spatial arrangement of the complex between eukaryotic initiation factor eIF-3 and 40 S ribosomal subunit. Cross-linking between factor and ribosomal proteins. *Biochim Biophys Acta 741*, 103-108.

Wilcox, S. A., Watson, J. M., Spencer, J. A., and Graves, J. A. (1996). Comparative mapping identifies the fusion point of an ancient mammalian X-autosomal rearrangement. *Genomics 35*, 66-70.

Wool, I. G. (1979). The structure and function of eukaryotic ribosomes. *Ann Rev Biochem 48*, 719-754.

Zinn, A. R., Alagappan, R., Brown, L., Wool, I., and Page, D. C. (1994). Structure and function of ribosomal protein S4 genes on the human and mouse sex chromosomes. *Mol Cell Biol 14*, 2485-2492.

Zinn, A. R., Bressler, S. L., Beer-Romero, P., Adler, D. A., Chapman, V. M., Page, D. C., and C.M., D. (1991). Inactivation of the Rps4 gene on the mouse X chromosome. *Genomics 11*, 1097-1101.

Chapter 4. *SMCX/SMCY*: Multiple, Independent Events

Background

Like *ZFX* and *RPS4X*, *SMCX* is ubiquitously expressed and escapes X inactivation and has a conserved Y homolog in humans (Agulnik et al., 1994; Wu et al., 1994). Unlike *ZFX* and *RPS4X*, *SMCX* also escapes X inactivation and has a conserved Y homolog in mice (Wu et al., 1994). Indeed, it is the only gene known in the sex-specific region of the mouse X chromosome (the part of the chromosome that does not recombine in male meiosis) that is known to escape X inactivation. In the human case, escape from X inactivation was demonstrated by the ananlysis of hamster-human hybrids containing either an active or inactive human X chromosome. In the case of mice, two alleles were found to be expressed in female mice with Searle's translocation [T(16;X)16H] that uniformly inactivated the intact X chromosome. *SMCX* represented the first case known in mice of a gene that is expressed from both inactive and active X chromosomes (Agulnik et al., 1994; Wu et al., 1994).

The function of *SMCX*, also known as *XE169*, and of its Y homolog are not known though SMCY protein has been defined as the male specific minor histocompatibility antigen H-Y in both humans and mice (Scott et al., 1995; Wang et al., 1995). Both *SMCX* and *SMCY* appear to be universally expressed; *SMCY* has been shown to be expressed in all adult and embryonic male tissues tested and in preimplantation embryos (Agulnik et al., 1994).

The case of *ZFX* suggests—and the case of *RPS4* does not contradict this—that rodents have the most terminally differentiated sex chromosomes among placental mammals. If that is so—and yet *SMCX/SMCY* remain even in rodents uninactivated/intact intimations of the sex chromosomes' shared ancestry—then no X inactivation of *SMCX* would be expected in any eutherians.

Methods

Extending 5' sequence of mouse Smcx. After mRNA was prepared from the kidney of a female Balb/c mouse (Hansen and Braman, 1992), cDNA was synthesized using Clontech's Marathon cDNA Amplification protocol; adaptors were ligated onto the ends of the cDNA. 5' RACE PCR was performed using the "Touchdown" conditions specified by Clontech, an adaptor primer and TAAACCTGAAGTTATCCACTTCCACGGC from mouse *Smcx*. Nested PCRs were performed on the resulting product with a second adaptor primer and CGGGTGGGCGGATCTTGCAAATGC from mouse *Smcx*. Products were cloned and sequenced as described in Chapter 3.

Methylation assay. Conditions were as described for *ZFX*. The primers used for *SMCX* were CCTCGGGCCCACCATGGAG and CTGATTTTCGCGATGTAGCC and amplify a 117 bp product that includes 3 CCGGs in human and 2 CCGGs in mice (figure 4.1).

Southern blotting. As described in Agulnik et al., 1994.

Screening of rabbit cDNA library. A cDNA library derived from the liver of an adult male rabbit was obtained from Clontech. It was treated like the dog library described in Chapter 3. The library was screened at low stringency as described in Chapter 3. The probe was a ^{32}P random primer labeled 370 bp fragment of mouse *Smcx* cDNA (pCM4 obtained from the Colin Bishop lab and described in Agulnik et al., 1994); PCR was performed with the primers CCTTCCAAGTTCAACAGTTATGG and CATACGTATGACTCAATAAACTGGG on pCM4 template. Conditions were as described for previous PCR reactions, but the annealing temperature was 60°C instead of 62°.

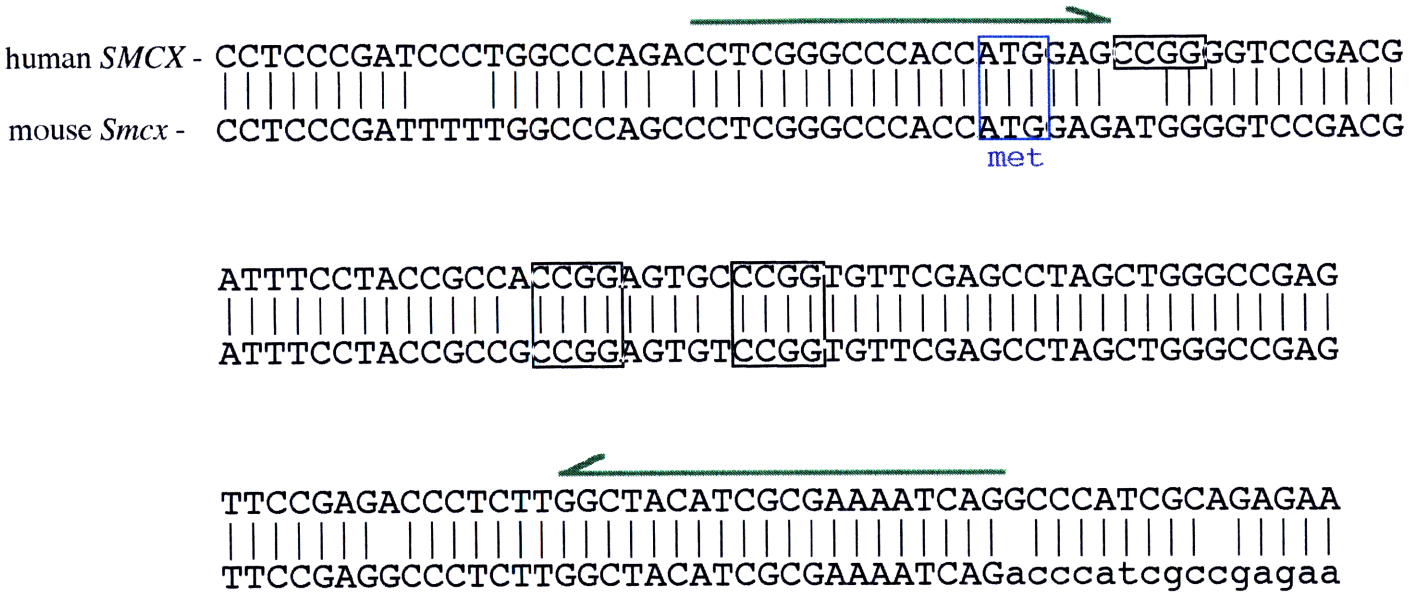


Figure 4.1 Alignment of human and mouse *SMCX* sequences in the 5' CpG island region. HpaII/MspI (CCGG) sites are shown in boxes. PCR primers used to assay methylation are outlined. The start codon is shown.

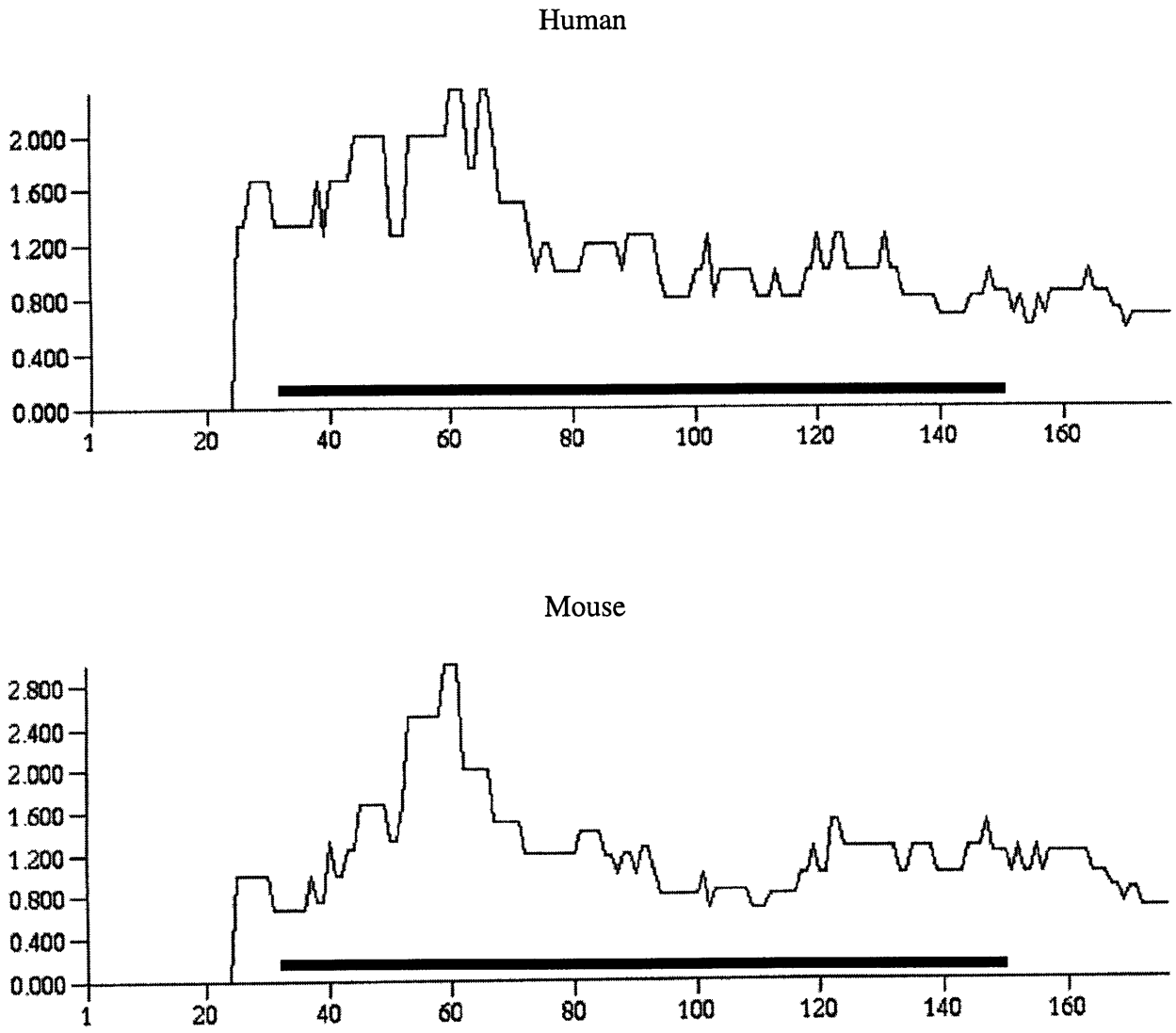


Figure 4.2 CpG to GpC dinucleotide ratios in the 5' regions of human and mouse *SMCX*, where the assay shown in figure 4.1 was designed. In mammalian genomic DNA the CpG to GpC ratio tends to be near 0.2 but in CpG islands approaches 1. Ratios were calculated in 50 bp windows. Areas where methylation was tested are marked with bars over the *x* axes.

Note that discrepancies are expected at the 5' and 3' ends of these kinds of graphs because dinucleotide counting occurs over windows. Irregularities are most pronounced when short sequences are analyzed.

Results

Full length human *SMCX* cDNA sequence that extended into the 5' UTR was available when this project began (Wu et al., 1994). Though much of the mouse *Smcx* cDNA sequence was published, it fell 100 bp short of the start codon and of the region into which this ubiquitously expressed gene's CpG island overlaps (Agulnik et al., 1994). I extended the available 5' mouse *Smcx* sequence by 5' RACE (Rapid Amplification of cDNA Ends). When 5' sequences from human and mouse *SMCX* were compared, conserved regions appeared among CpG island-like sequences (figure 4.2) that included testable restriction enzyme sites (figure 4.1). If an assay is conserved between humans and mice, which are about as diverged from each other as any two placental mammals, it is likely to apply to a wide array of eutherians. Indeed, the *SMCX* methylation assay designed to apply to both humans and mice, could be extended to all other tested eutherians. That is, a PCR product of the expected size could be amplified using any eutherian DNA as template, and the product could be eliminated when template DNA was first digested with MspI.

Instead of finding that *SMCX* universally escapes X inactivation, 5 of 11 tested eutherian orders (6 of 18 species) showed methylation in females (but not in any males) (figure 4.3). *SMCX* appears to be subject to X inactivation in lagomorphs, caviomorphs, artiodactyls (at least cow and goat), edentates, and insectivores (figure 4.4). These orders cannot be grouped together phylogenetically to exclude those that exhibit no methylation and presumably escape X inactivation. For instance, cetaceans and artiodactyls constitute a clade, yet whales show no *SMCX* X inactivation while goats and cows (artiodactyls) do exhibit X inactivation of *SMCX*.

Just as Y homologs of *ZFX* show up by Southern blotting, *SMCY* homologs can be detected as male-specific Southern blot bands upon hybridization with mouse *Smcx* cDNA. Agulnik et al. observed male-specific bands in human, mouse, rabbit, pig, horse,

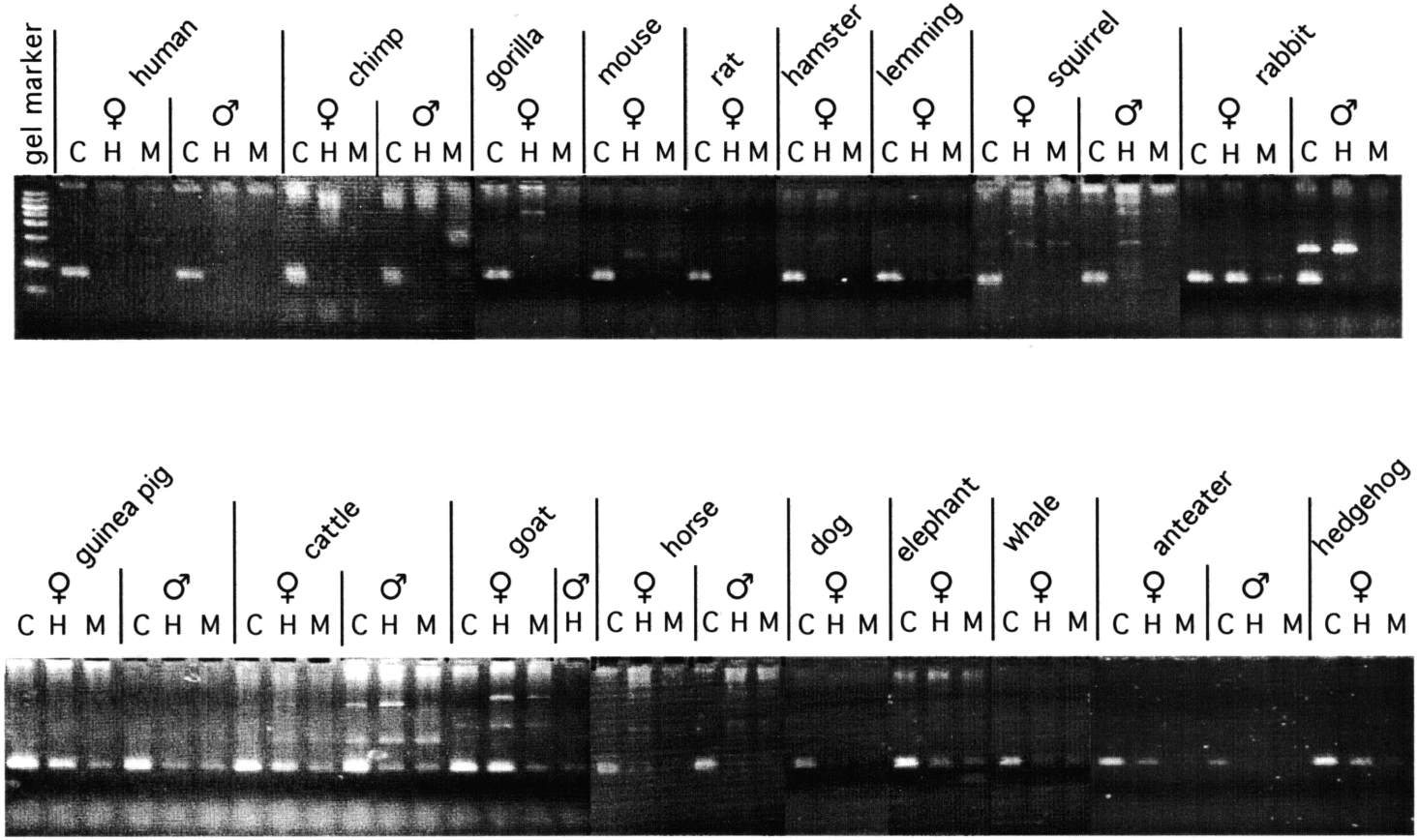


Figure 4.3 *SMCX* methylation results. Female and male genomic DNAs digested with HpaII (H) or MspI (M) or HindIII (C, "control") were used as alternate templates in PCR assay diagrammed in figure 4.1. PCR products were visualized by ethidium bromide/UV staining after agarose gel electrophoresis.

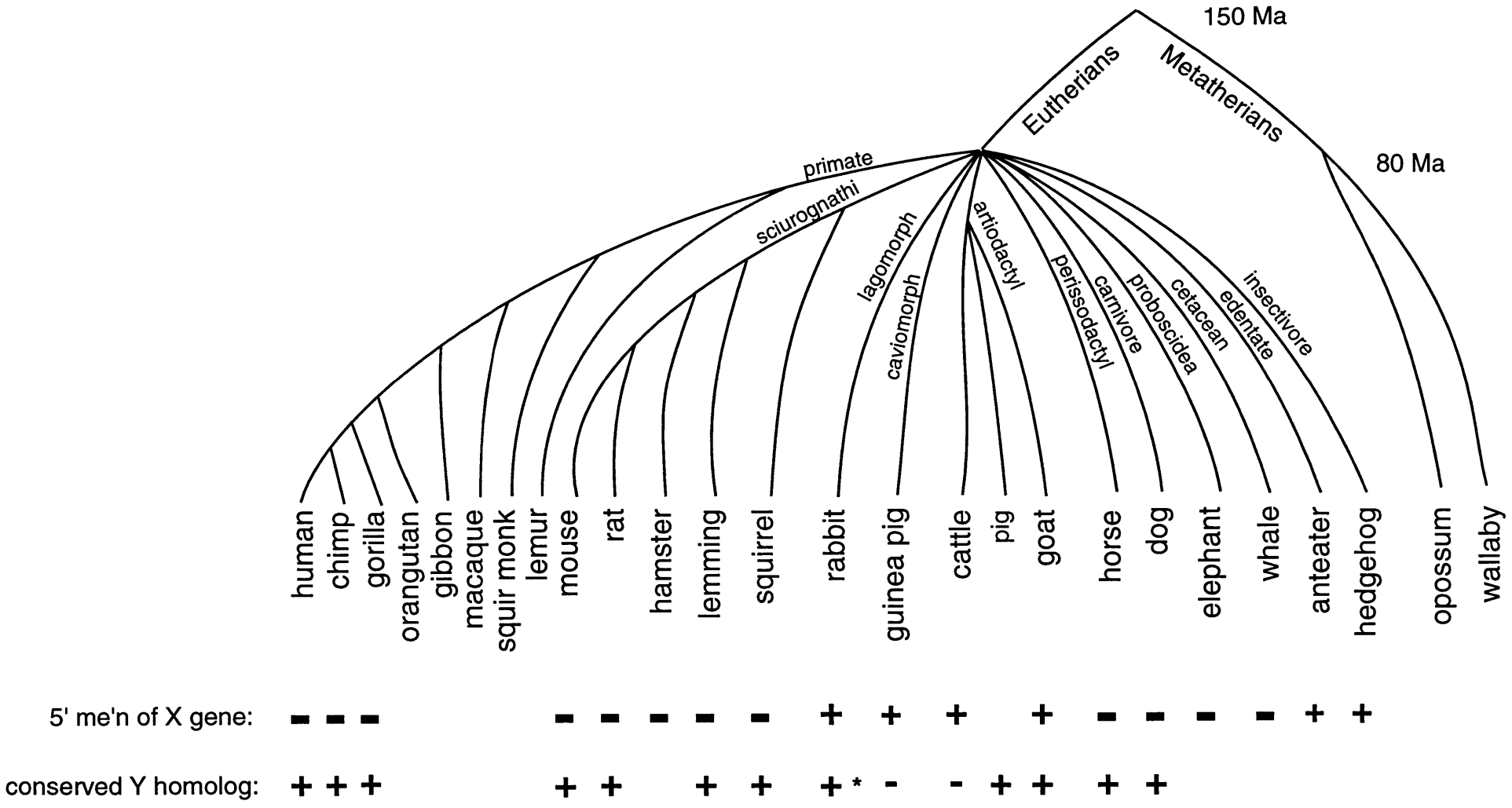


Figure 4.4 Results of *SMCX* methylation and Y homolog studies in mammalian species, arranged phylogenetically. The presence of a Y homolog was judged by Southern blotting. *SMCY* homolog results were obtained with a probe derived from the *Smcx* coding region, pCM4. Agulnik et al reported most of the Y homolog results in 1994.

* In the case of rabbit, though a male-specific band was observed by Southern blotting, male cDNA library screening revealed but one species of *SMCX* homologous sequence. Thus, though sequences homologous to *SMCX* still exist on the rabbit Y chromosome, an *SMCY* gene apparently is no longer expressed.

dog, and kangaroo but not in cattle (Agulnik et al., 1994). I confirmed the previous findings for the species shown in figure 4.4 and extended them: Southern hybridization also revealed male-specific bands for more primates (chimp, gorilla) and more rodents (rat, lemming, squirrel); no male band appeared in the case of guinea pig, a caviomorph. The species which lacked any apparent *SMCY* upon Southern hybridization, guinea pig and cattle, also exhibited *SMCX* X inactivation. Most of the eutherians which have a Y homolog of *SMCX*—primates, rodents, horse and dog—also show no methylation or X inactivation of *SMCX*. But two species seem at odds with the usual correlation of X inactivation and Y degeneration. Rabbit and goat, members of two different eutherian orders showed male-specific Southern bands even though the X-linked genes appear to be subject to X inactivation.

Still, it is conceivable that though a Y-specific band is observed on a Southern, no Y homolog is expressed; sequence conservation may linger in the genome even after a gene ceases to be expressed. I examined rabbit in greater detail to see whether the Y homologous sequences detected by Southern hybridization represent an expressed *SMCY* or merely the remains of a gene still capable of Southern hybridization but no longer expressed widely now that *SMCX* in rabbit appears subject to X inactivation.

A male rabbit liver cDNA library was screened at low stringency with a small (370 bp) portion of the mouse *Smcx* cDNA. Thirteen overlapping clones were found that aligned with known *SMCX* sequences, and all represented one gene. If two different genes existed, they would not have been hard to distinguish. For example, the X and Y forms at the nucleotide level are 26% diverged in mouse and 21% diverged in human (Agulnik et al., 1997). In humans and mice, *SMCX* and *SMCY* are roughly evenly expressed (Agulnik et al., 1994), so finding 13 clones all representing just one gene strongly suggests that only an *SMCX* (an SMCX subject to X inactivation) exists in rabbit and that the male-specific Southern blot band represents an *SMCY* gene that is

likely degenerating in sequence and is either no longer active at all or no longer as widely expressed as the X homolog.

Discussion

The case of *SMCX* shows that the myomorph sex chromosomes do not represent the most highly evolved sex chromosomes. No eutherian can be said to have the most highly evolved sex chromosomes. Genes can be found in all tested eutherians that still escape X inactivation and still retain conserved Y homologs, yet different sets of genes show these properties among different placental mammals. The accumulating case studies suggest that evolution of the sex chromosomes occurs gene by gene, with one X-Y pair dying in one lineage and another X-Y pair dying independently in another, in the general tendency toward greater X dosage compensation and Y degeneration. That *SMCX* must have acquired X inactivation more than once to account for the pattern of methylation observed in the tested placental mammals underscores how X inactivation can apparently be acquired piecemeal.

Rabbit, and perhaps goat, show telling transition states in the evolution of X-Y homologous genes. The Y homolog in rabbit appears to be newly diverged, still conserved enough at the level of sequence to be detected by hybridization but no longer functionally interchangeable with the X homolog which has acquired X inactivation. What seemed at first a paradox in rabbit turns out not to be so strange. With *SMCX* X inactivated and the Y homolog no longer (at least widely) expressed, only one copy of *SMCX* will be expressed in both female and male cells. X inactivation has not been acquired without the corresponding degeneration of the Y homolog.

References

Agulnik, A. I., Bishop, C. E., Lerner, J. L., Agulnik, S. I., and Solovyev, V. V. (1997). Analysis of mutation rates in the *SMCY/SMCX* genes shows that mammalian evolution is male driven. *Mamm Genome 8*, 134-138.

Agulnik, A. I., Mitchell, M. J., Lerner, J. L., Woods, D. R., and Bishop, C. E. (1994). A mouse Y chromosome gene encoded by a region essential for spermatogenesis and expression of male-specific minor histocompatibility antigens. *Hum Molec Genet 3*, 873-878.

Agulnik, A. I., Mitchell, M. J., Mattei, M. G., Borsani, G., Avner, P. A., Lerner, J. L., and Bishop, C. E. (1994). A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. *Hum Molec Genet 3*, 879-884.

Hansen, C., and Braman, J. (1992). Isolation of mRNA. *Strategies 5*, 66-67.

Scott, D. M., Ehrmann, I. E., Ellis, P. S., Bishop, C. E., Agulnik, A. I., Simpson, E., and Mitchell, M. J. (1995). Identification of a mouse male-specific transplantation antigen, H-Y. *Nature 376*, 695-698.

Wang, W., Meadows, L. R., den Haan, J. M., Sherman, N. E., Chen, Y., Blokland, E., Shabanowitz, J., Agulnik, A. I., Hendrickson, R. C., Bishop, C. E., and et al. (1995). Human H-Y: a male-specific histocompatibility antigen derived from the SMCY protein. *Science 269*, 1588-1590.

Wu, J., Ellison, J., Salido, E., Yen, P., Mohandas, T., and Shapiro, L. J. (1994). Isolation and characterization of XE169, a novel human gene that escapes X-inactivation. *Hum Molec Genet 3*, 153-160.

Wu, J., Salido, E. C., Yen, P. H., Mohandas, T. K., Heng, H. H., Tsui, L. C., Park, J., Chapman, V. M., and Shapiro, L. J. (1994). The murine Xe169 gene escapes X-inactivation like its human homologue. *Nat Genet 7*, 491-496.

Chapter 5. *ALD*: A Control

Background

The relationship between CpG island methylation and X inactivation has been most thoroughly examined in humans and mice, species long and comprehensively studied by molecular biologists. Properties shared by these two species can reasonably be extrapolated to other placental mammals. Using CpG island methylation as a marker of X inactivation in a broad range of eutherians seemed sensible. Still, direct evidence that this is an appropriate strategy was sought. CpG island methylation is not associated with X inactivation in tested marsupials (Kaslow and Migeon, 1987), so this mechanism is not conserved among all mammals, but is it conserved among eutherian mammals? In studying *ZFX*, for example, we concluded that the absence of methylation in all eutherians except myomorphs suggests that the gene escapes X inactivation in all but myomorphs, but this conclusion would be invalid if other eutherians do not exhibit CpG island methylation associated with X inactivation.

A gene was sought that is subject to X inactivation in the species in which its expression has been examined and that has a 5' CpG island conserved enough for the widespread application of a single methylation assay, as in the case of *ZFX* and *SMCX*. Over 150 X-linked genes' 5' sequences were compared between humans and rodents before one was found that fit these criteria, the gene *ALD*. Deficiency of *ALD* causes X-linked recessive inheritance of adrenoleukodystrophy (hence the name of the gene), a disorder with neurological and endocrinologic symptoms (Aguilar et al., 1967). Two types of clones were found in heterozygous humans, demonstrating X inactivation of the gene. That is, when cells were cultured from a female with one sound and one defective allele of *ALD*, some lineages of cells were found to express *ALD* protein while other lineages expressed none (Migeon et al., 1981).

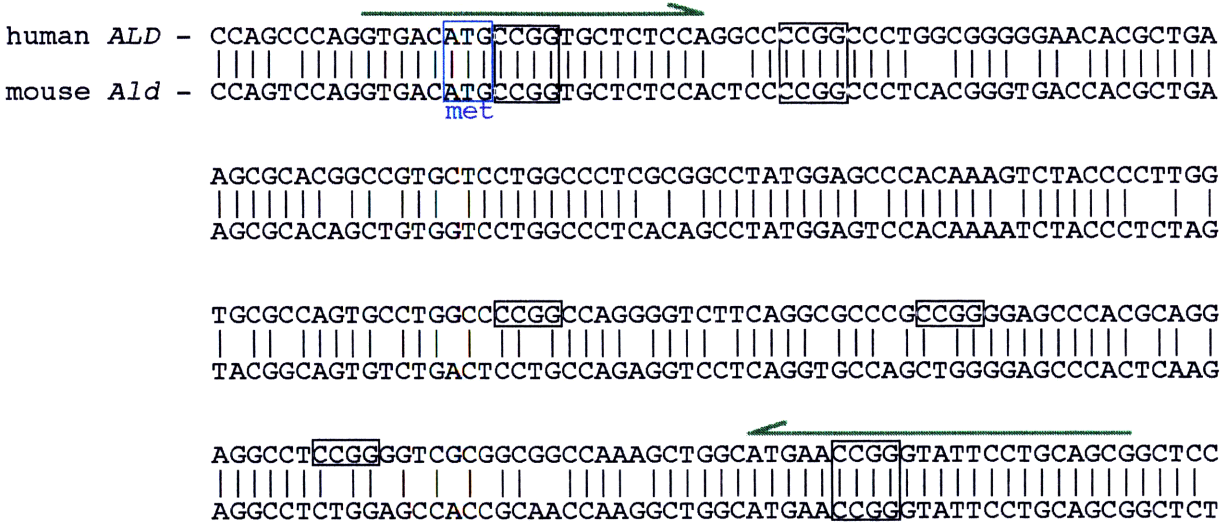


Figure 5.1 Alignment of human and mouse 5' CpG island-like *ALD* cDNA sequences. HpaII/MspI (CCGG) sites, PCR primers used to assay methylation, and the start codon are shown.

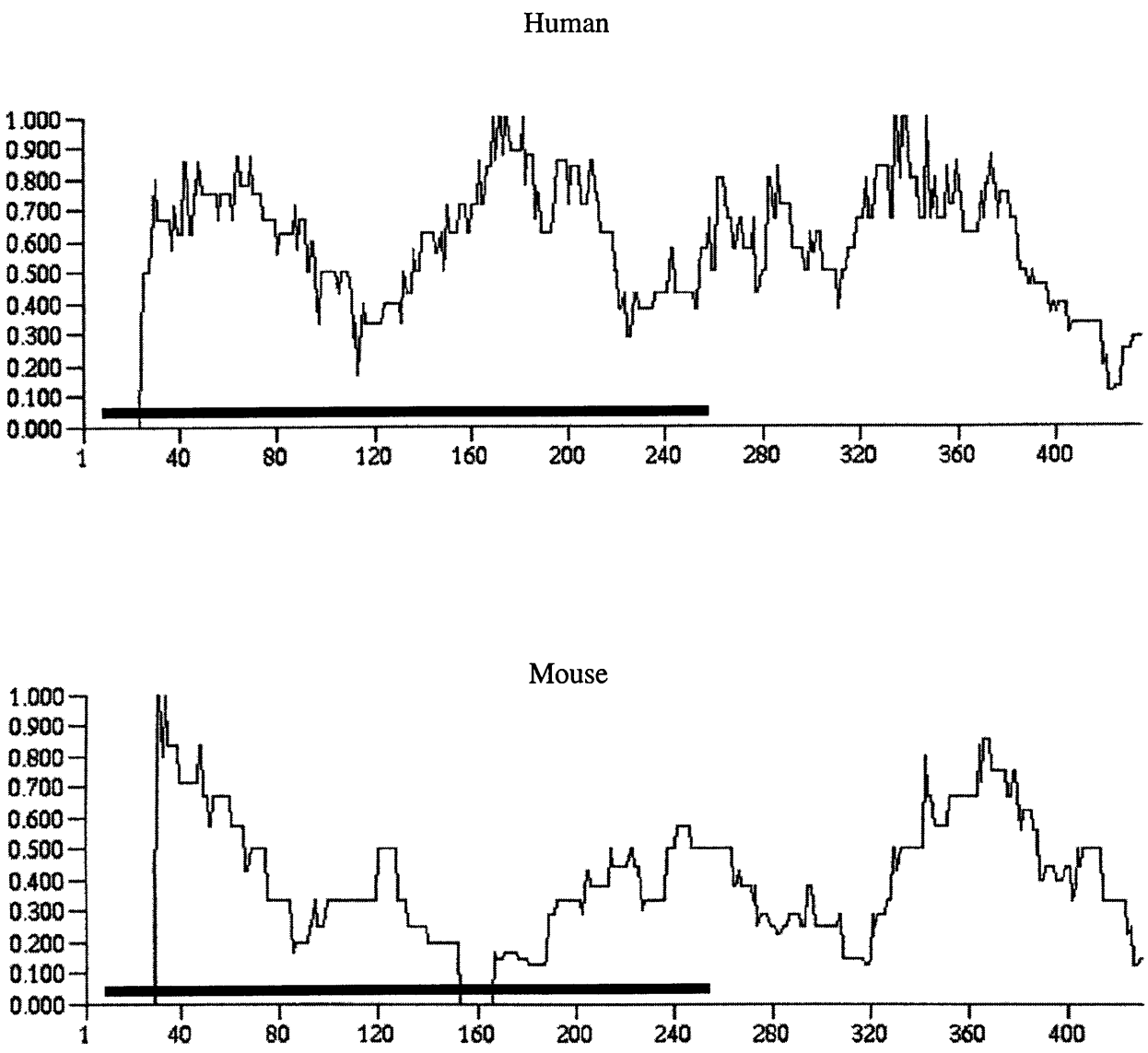


Figure 5.2 CpG to GpC dinucleotide ratios in the 5' regions of human and mouse *ALD*, where the assay shown in figure 5.1 was designed. In mammalian genomic DNA the CpG to GpC ratio tends to be near 0.2 but in CpG islands approaches 1. Ratios were calculated over 50 and 60 bp windows. Areas where methylation was tested are marked with bars over the *x* axes.

Note that discrepancies are expected at the 5' and 3' ends of these kinds of graphs because dinucleotide counting occurs over windows.

Methods

ALD Methylation assay. Conditions were as described for *ZFX*, except that a PCR annealing temperature of 60°C instead of 62° was used, and 0.5 μL of DMSO was added to each PCR reaction. The PCR primers used were GTGACATGCCGGTGCTCTCCA and CGCTGCAGGAATACCCGGTTCAT and amplify a 226 bp product including 6 HpaII/MspI sites in humans and 3 such sites in mouse (figure 5.1). The site of this assay has CpG island character (figure 5.2)

Results

The *ALD* 5' PCR assay, designed to amplify products from both humans and mice, also amplified product of the expected size from most eutherians, a product that disappeared if the template had been previously digested with MspI, showing that CCGG restriction enzyme sites are also conserved in the tested mammals. No PCR amplification could be obtained from the tested artiodactyls and cetaceans; perhaps these two orders, which constitute a clade, share a sequence divergence that prevents the use of the PCR assay.

In every tested placental mammal—17 species representing 9 Orders—female DNA exhibited methylation while male DNA, where the single copy of *ALD* will not be inactivated, showed no methylation (figures 5.3 and 5.4).

Conclusion

The case of *ALD* provides an experimental demonstration that CpG island methylation should serve as a marker of X inactivation widely among eutherians.

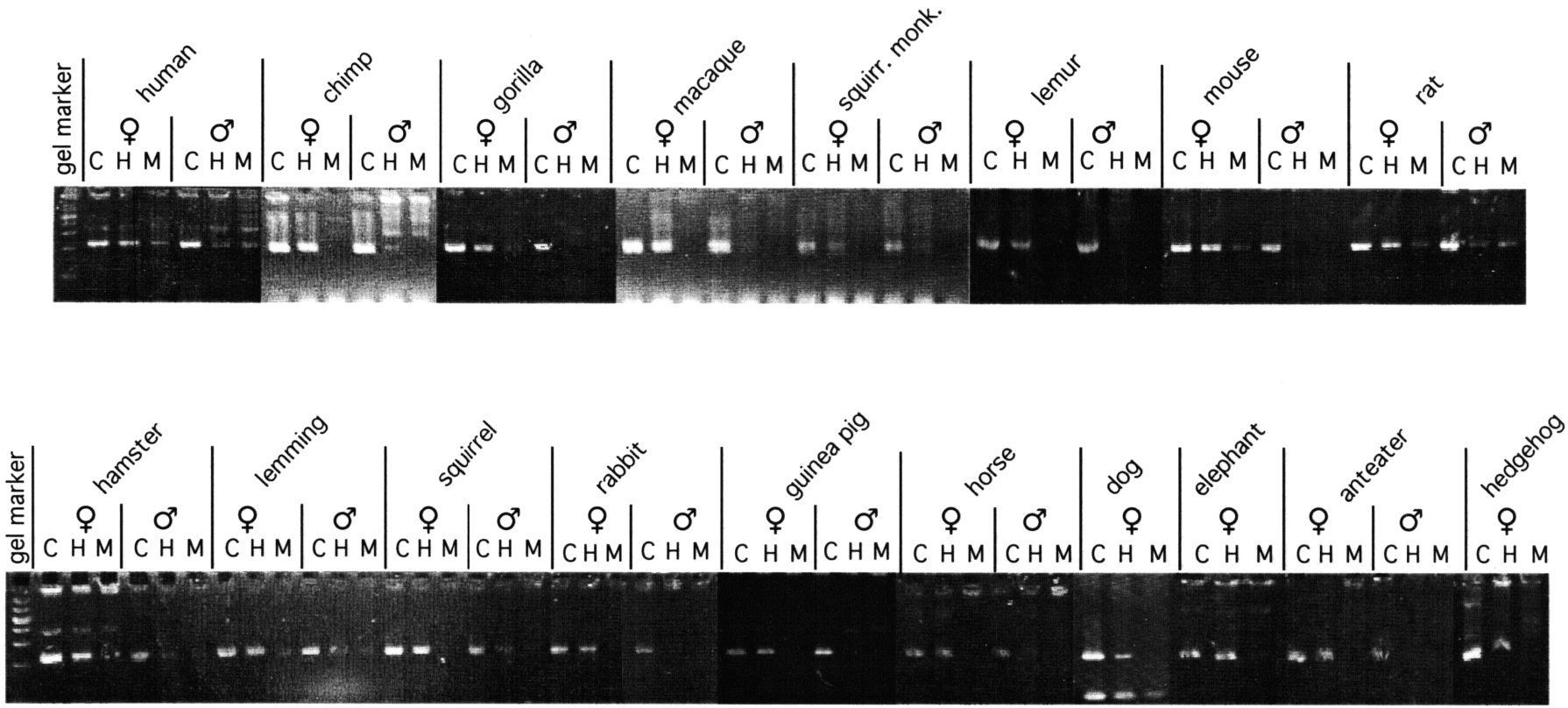


Figure 5.3 *ALD* methylation results. Female and male genomic DNA digested with HpaII (H) or MspI (M) or HindIII (C, "control") were used as alternate templates in PCR assay diagrammed in figure 5.1. PCR products were visualized by ethidium bromide/UV staining after agarose gel electrophoresis.

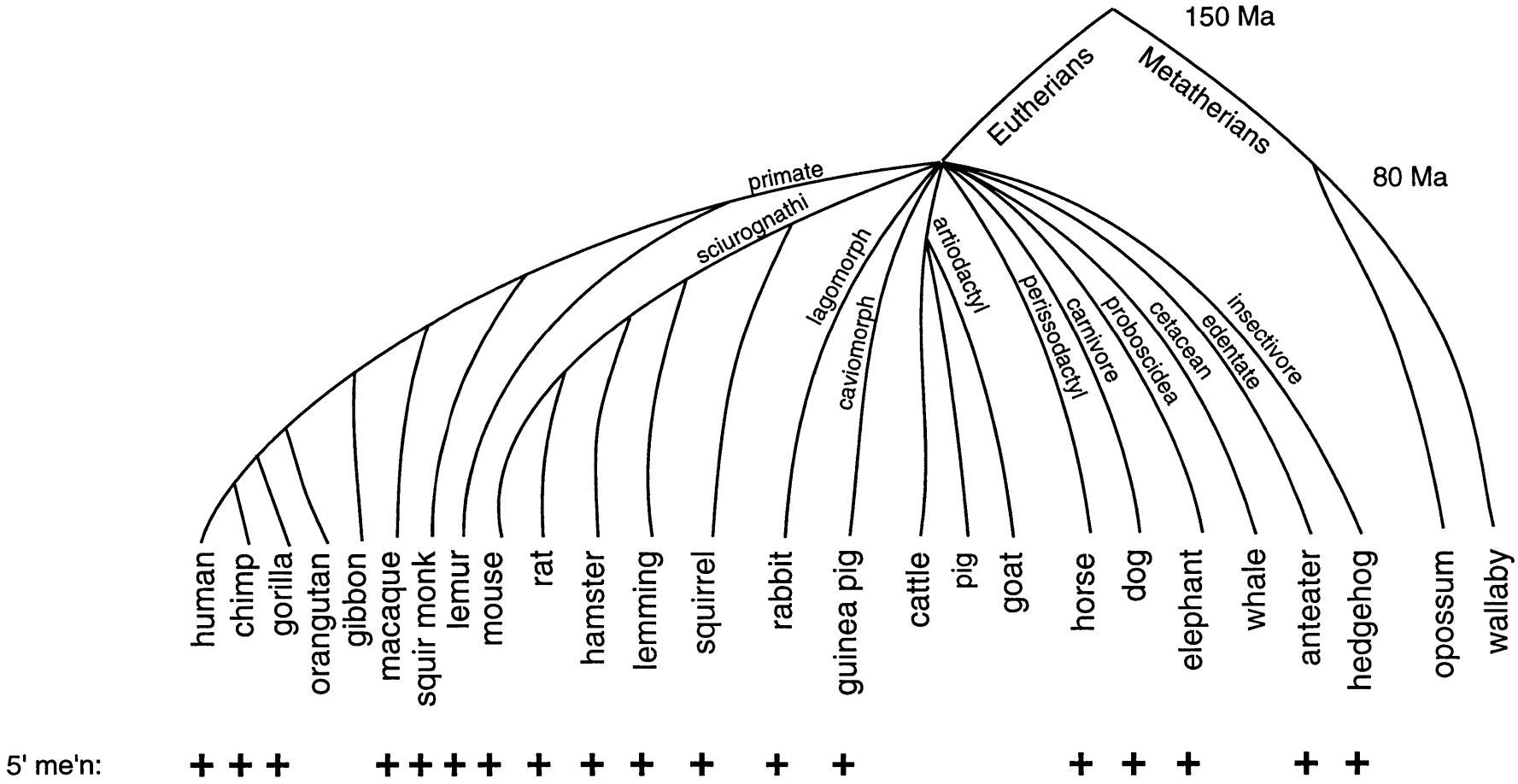


Figure 5.4 Results of *ALD* methylation studies in mammalian species, arranged phylogenetically.

References

Aguilar, M. J., O'Brien, J. S., and Taber, P. (1967). *The syndrome of familial leukodystrophy, adrenal insufficiency and cutaneous melanosis.* In Inborn Disorders of Sphingolipid Metabolism, S. M. Aronson and B. W. Volk, eds. (Oxford: Pergamon Press), pp. 149-166.

Kaslow, D. C., and Migeon, B. R. (1987). DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation. *Proc Natl Acad Sci USA 84*, 6210-6214.

Migeon, B. R., Moser, H. W., Moser, A. B., Axelman, J., Sillence, D., and Norum, R. A. (1981). Adrenoleukodystrophy: evidence for X linkage, inactivation, and selection favoring the mutant allele in heterozygous cells. *Proc Natl Acad Sci USA 78*, 5066-5070.

Chapter 6. Discussion

The Dynamism of Sex Chromosomes

When genes escaping X inactivation were first discovered, it was not clear what they represented. The finding that almost all of the genes which escaped X inactivation in humans are subject to X inactivation in mice raised the possibility that the cases of escape from X inactivation arose as specialized, exceptional events. Examination here of escape from X inactivation more widely among placental mammals suggests that the instances of escape from X inactivation represent a primitive condition, found in the ancestor of eutherian mammals. The genes which resist inactivation represent transition states, pockets of incompletion, in the differentiation of the sex chromosomes.

The genes that escape X inactivation, most of which have conserved Y homologs, appear to be evolutionary holdovers, intimating the histories of the X and Y chromosomes. Two principal tendencies—Y degeneration and the acquisition of X dosage compensation—characterize the evolution of the sex chromosomes and drive the differentiation of the X and the Y from each other. Even in mammalian sex chromosomes, which are often characterized as terminally differentiated (Bull, 1983), these tendencies still hold sway. The essential forces that have driven sex chromosome differentiation in mammals for millenia still impel the expansion of dosage compensation on the X and the decay of genes on the Y.

The mammalian X and Y chromosomes are thought to have originated as an unremarkable pair of autosomes, when an allele arose on the Y that affected sex determination. Since then, the two chromosomes have been diverging from each other. The lack of recombination along most of the length of the Y is thought to have driven or at least enabled the decay of the Y chromosome. To equalize dosage with males, who only retain one copy of the X and the genes on it inherited from the long-ago autosome, females acquired a mechanism of dosage compensation to limit the expression of the two X chromosomes to the equivalent of the one X chromosome in males. If the dosage

of a gene on the sex chromosomes matters, two strategies serve equally well to balance expression in males and females. Either a gene can retain a functionally conserved Y homolog and not be subject to X inactivation, in which case both males and females will express two copies of the gene per cell, or a gene can degenerate on the Y and become subject to X inactivation, in which case both males and females will express one copy of the gene per cell, only from the active X.

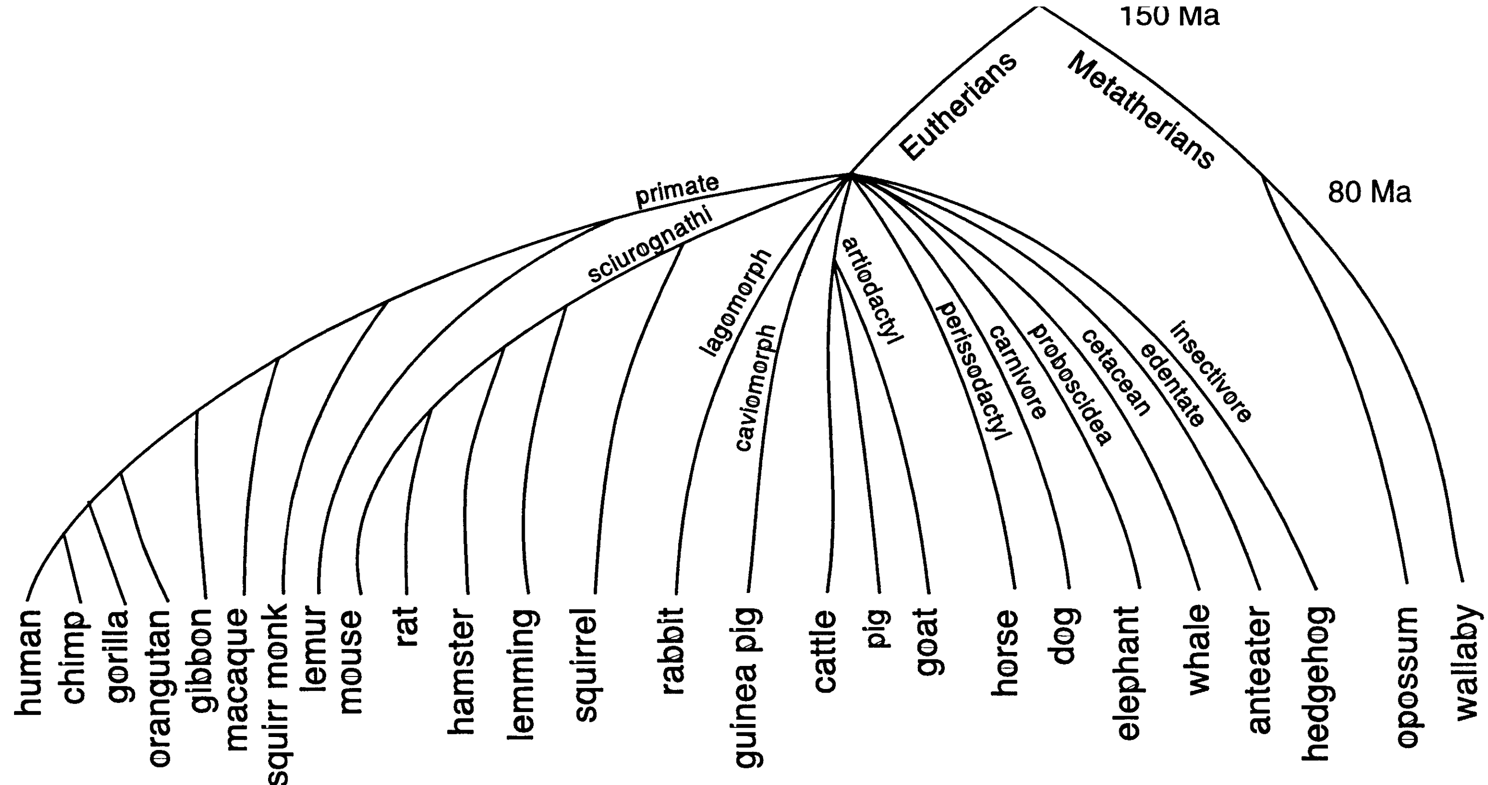
The work described in the previous chapters, particularly the study of *ZFX* and *ZFY*, indicates that expressing two copies of sex chromosome genes per cell is the more original, the ancestral condition while expressing just one copy of sex chromosome genes per cell represents a derived state. Figure 6.1 shows an overview of the results reported throughout this text.

Yet X inactivation and Y degeneration, though logically linked, are not mechanistically linked processes. A gene on the X is not likely to acquire X inactivation at just the moment that its Y homolog ceases to be expressed. What is the temporal relationship between X inactivation and Y degeneration?

The evidence suggests that Y degeneration, or at least divergence, precedes and powers the acquisition of X inactivation. No cases are known where a gene is subject to X inactivation even though a functionally conserved homolog exists on the Y chromosome. Several cases are now known in which a gene continues to evade X inactivation though the Y homolog is no longer expressed or is expressed in only a limited way, perhaps serving a diverged function. Such cases imply that the acquisition of X inactivation can lag behind the loss of a co-functional homolog. The following human genes lack Y homologs or have Y homologs that have degenerated into unexpressed pseudogenes though they (still) escape X inactivation: *GS1, STS, KAL1, PCTK1, UBE1, TIMP1, SB1.8* (figure 1.1).

Figure 6.1 Results of *ZFX*, *RPS4*, *SMCX*, *ALD*, *DBX,* and *IGF2R* methylation and Y homolog studies in mammalian species, arranged phylogenetically. Methylation data are shown in paranthesis for homologs tested in marsupials, where X inactivation does not involve 5' CpG island methylation.

* When Y homologs of *SMCX* were sought by Southern hybridization, rabbit had a male-specific band. But cDNA library screening revealed but one species of *SMCX* homologous sequence (Chapter 4).



	human	chimp	gorilla	orangutan	gibbon	macaque	squirr monk	lemur	mouse	rat	hamster	lemming	squirrel	rabbit	guinea pig	cattle	pig	goat	horse	dog	elephant	whale	anteater	hedgehog	opossum	wallaby
5' me'n of X gene:																										
ZFX:	-	-	-					-	+	+	+	+	-	-	-	-		-	-	-	-	-	-	-	(-)	(-)
RPS4:	-	-	-	-	-	-	-		+	+		+	+	+	+	*no sites*		*no sites*	+	+		+	+	+		
SMCX:	-	-	-						-	-	-	-	-	+	+	+		+	-	-	-	-	+	+		
ALD:	+	+	+			+	+	+	+	+	+	+	+	+	+				+	+	+		+	+		
DBX:	-								-		-	-		-		-		-			-	-	-	-	(-)	
5' me'n of autosomal gene: ***IGF2R:***	-								+	+																
conserved Y homolog:																										
ZFX:	+	+	+						-	-		-	+	+	+	+		+	+	+						
RPS4:	+								-	-				-		-				-					+	
SMCX:	+	+	+						+	+		+	+	+*	-	-	+	+	+	+						

The case of rabbit *SMCX/SMCY* described in Chapter 4 is another case that suggests that X inactivation has not occurred without Y degeneration. The rabbit *SMCX* appears to be subject to X inactivation. Even though the remains of a Y homolog can be observed on a Southern blot, on further examination that gene appears to be dead or diverged. Either it is no longer expressed, or it is not expressed ubiquitously. The lingering sequence homology on the Y suggests the gene has only recently ceased to be expressed or diverged in function.

UBE1 is another illustrative case. On the surface, it appears to contradict the normal dosage arguments: the X-linked gene escapes inactivation in humans though no human Y homolog can be found while mouse *Ube1x* is subject to X inactivation though a mouse Y homolog exists. This state of affairs would seem to lead to uneven dosage in males and females in both mice and humans (1 copy/cell in human males and mouse females; 2 copies/cell in human females and mouse males). But on closer inspection, the situation is not so paradoxical. The human case is not unusual: probably a case of X inactivation lagging behind Y degeneration. And the seemingly contradictory case in mouse no longer seems strange when one realizes that mouse *Ube1y* is solely expressed in the testis (Odorisio et al., 1996), apparently having gone off on a functional fork and acquired a novel, male-specific function.

No counterexamples seem to dispute the conclusion that Y degeneration or divergence seems to precede X inactivation. No case is known where a Y homolog persists that is expressed in an equivalent way to an X-inactivated homolog though several cases exist where a gene escapes X inactivation even though the Y homolog has degenerated, implying that Y degeneration precedes the acquisition of X inactivation. In cases where a Y homolog persists though the X gene is inactivated (e.g. the mouse *Zfy* genes or mouse *Ube1y*), the Y genes have different, limited expression patterns, apparently new male-specific roles compared to the X homologs. So, to be precise, it is

not necessary that a Y gene die to drive the acquisition of X inactivation; it must lose the ability to function interchangeably with the X homolog whether by ceasing to be expressed or diverging in its functional role.

The gene-by-gene acquisition of X inactivation in eutherian mammals seen in the data reported here seems to me to support the notion that X inactivation is not the driving force to Y degeneration but instead occurs in response to Y degeneration (or specialization). If sweeps of X inactivation had been observed instead that suddenly made whole sets of genes subject to X inactivation in one lineage, the mechanistic hurdle to acquisition of X inactivation would have seemed more momentous. As it is, X inactivation can apparently be acquired in a spotty way, gene by gene.

How can one explain the apparent chromosome-residence effects on the expression of some genes that escape X inactivation, where the allele on the inactive X is expressed at lower rates compared to the allele on the active X? Such an effect was observed for mouse *Smcx* recently (Sheardown, et al., 1996; Carrel et al., 1996) and human *STS* fifteen years ago (Migeon, et al., 1982). One could speculate that chromatin effects on genes that reside on the the inactive X somehow dampen the expression of these genes, even if they resist outright inactivation. But suppose, instead, that diminished expression from X_i and enhanced expression from X_a reflect the evolutionary stage of these genes. Perhaps the genes are on their way to acquiring X inactivation because their Y homologs are beginning to diverge in function. Indeed, in humans, the Y homolog of *STS* is no longer expressed, a slowly decaying pseudogene. Perhaps further study of mouse *Smcy* will show that it is beginning to diverge in function compared to *Smcx*.

Saying that the acquisition of X inactivation does not have to occur in massive sweeps does not mean that changes in X inactivation status occur haphazardly. Though the patterns of gene activity compiled in figure 6.1 are unique for each gene analyzed, in phylogenetic context the results are nonrandom for each gene. The acquisition of X

inactivation apparently occurs rarely enough that observed patterns of methylation/X inactivation make sense in phylogenetic context. For example, the primates cluster together, the myomorphs cluster together; *RPS4* appears to escape X inactivation in all primates though in no other tested eutherians; *ZFX* is subject to X inactivation and *SMCX* escapes X inactivation in all tested myomorphs.

Two overarching tendencies characterize the history of the sex chromosomes: a tendency for Y genes to decay (or at least diverge) and a tendency for the range of X dosage compensation to increase. The conclusions reported here are perhaps most notable for what they reveal about the dynamism of the sex chromosomes. Even in the recent past, among eutherian mammals with their "terminally differentiated" X and Y chromosomes, the sex chromsomes have continued to evolve, and the particulars of how they have evolved seem idiosyncratic. Twists of fate appear to have led to the X inactivation of one subset of genes in one lineage and to the X inactivation of altogether different subsets in other lineages.

Figure 1.1 shows the pronounced clustering of genes that escape X inactivation on the short arm of the human X chromosome. Several genes on the short arm of the human X map to autosomal sites in marsupials and monotremes (Graves, 1987; Graves and Watson, 1991). The marsupials and monotremes appear to have diverged independently from the placental mammalian lineage, the monotreme line having diverged from the masupial-placental line about 40 million years before the marsupials diverged from the placental mammalian lineage. An X-conserved region, comprising at least a portion of the X in all tested mammals, including marsupials and monotremes, represents the original mammalian X that has remained intact over millenia. The genes *UBE1*, *SMCX*, and *RPS4X* are among those with copies on the X and Y chromosomes of marsupials as well as eutherians; these X genes and their Y homologs must have existed in distinct X and Y forms for at least 150 million years, since the metatherians (marsupials) diverged from the eutherians.

How did most of the human Xp genes, many of which retain Y homologs, wind up on the X chromosome? Jenny Graves's addition-attrition hypothesis of sex chromosome evolution supports the notion of the dynamic condition of sex chromosomes (Graves, 1995). She postulated that, in cycles, portions of autosomes transpose onto the pseudoautsomal regions of one of the sex chromosomes, that these autosomal regions recombine onto the other sex chromosome. The expanded pseudoautosomal regions (PARs) of the sex chromosomes are then subject to the pressures of sex chromosome evolution. Attrition of the Y drives the increasing sway of X inactivation. Instances of gene duplication, for example of steroid sulfatase genes, have been documented in the ancestral PAR (Meroni et al., 1996). Little concordance in gene content has been observed between the mouse and human pseudoautosomal regions suggesting that these regions have changed much during recent mammalian evolution. For instance, *Sts* is the only gene known in the mouse PAR, but in humans *STS* is not in the PAR but just outside it; *STS* escapes X inactivation but its Y homolog has degenerated into a pseudogene. Probably *STS* was pseudoautosomal in the eutherian ancestor shared by humans and mice, has remained pseudoautosomal in mouse, but has been rearranged out of the PAR in humans.

Why have some gene pairs persisted so long on the Y and remained resistant to X inactivation, like *RPS4X/RPS4Y* and *SMCX/SMCY*, while others that were pseudoautosomal not long ago have acquired X inactivation and quickly lost their Y homologs? The escape from X inactivation of many genes, especially those on Xp and the persistence of their Y homologs may reflect their recent acquisition into the sex-specific portions of the X and Y. But those gene pairs that resisted Y degeneration and X inactivation though they have existed on the sex chromosomes for over a hundred million years must, it seems to me, have lasted so long because their dosage is critical. The period of discrepancy in dosage between the sexes that would result from limiting expression from the Y without accompanying X inactivation of the homolog (or, less

likely, from X inactivating a gene while retaining a Y homolog) might be difficult to tolerate.

Turner Syndrome (TS) is a tangible representation of the importance of dosage of at least some of the genes that escape X inactivation and retain Y homologs. Haploinsufficiency of X-Y homologous genes is likely to cause TS (Ferguson-Smith, 1965). Genes which have persisted long as X-Y homologous pairs in humans seem likely to be the strongest TS candidates. Evolutionary studies may also help direct the search for genes important in the etiology of TS. Genes which have not become subject to X inactivation in any or most mammalian lineages—like *DBX* (see Appendix)—may be particularly good candidates for TS since their dosage seems likely to be finely tuned, with a high threshold for any adjustment of dosage.

Figure 6.1, with its accumulation of data, begs a reconsideration of cladograms. In Chapter 1, I discussed how cladograms for mammals have been drawn and in chapters since have settled on a conservative depiction of eutherian phylogeny—one that does not attempt to refine the sequence of radiation among the eutherian orders. My strategy has been to overlay my data onto trees drawn on the basis of other data, as a way to interpret the history of X-Y homologous genes rather than as a means for coming up with alternative cladograms. But the data I gathered could itself be used as the basis for drawing a cladogram. The several types of binary data shown in figure 6.1 can be used to array the mammals examined in a way that would require invoking the fewest number of molecular changes. Based purely on the data shown in figure 6.1, the most parsimonious cladogram would show the following order of branching: after marsupials, the primates (note the *RPS4* data); then the rodents, perissodactyls, carnivores, proboscidea and cetaceans would all diverge from the eutherian lineage before the lagomorphs, caviomorphs, artiodactyls, edentates and insectivores diverged from each other (note the *SMCX* data).

Reliable phylogenetic trees are seldom drawn from consideration of just a few characters, and a tree drawn on the basis of the data in figure 6.1 is not likely to supplant trees based on extensive sequence comparisons. For example, the tree I describe places whales far from artiodactyls, but cetaceans appear to be conclusively related to artiodactyls on the basis of other data (see chapter 1). In all, the tree I describe looks little like the tree shown in figure 1.4 B, which was drawn based on the sequence analyses of others and, because based on more data, seems more likely to be reliable. Perhaps the most enticing claim the tree I describe makes is that primates diverged early among eutherian orders; this would neatly account for the X inactivation, Y homolog data observed for *RPS4X* and this positioning of primates, though not supported by other literature, does not appear to contradict any firmly established phylogenetic evidence.

The gene-by-gene acquisition of X inactivation described in this text relates to considerations of the mechanism of X inactivation. The mechanism of X inactivation is still mysterious, but it appears to be a sweeping, chromosome-directed process: the whole chromosome seems morphologically, behaviorally, and biochemically distinctive. Cytogenetically, the inactivated X is heterochromatinized, observed as a Barr body (Walker et al., 1991); it replicates late (Grumbach et al., 1963); it lacks histone H4 acetylation (Jeppesen and Turner, 1993) and is coated with *XIST* RNA (Clemson et al., 1996); the *XIST* gene must be contiguous to chromosome regions that are inactivated. X inactivation looks like a powerful cis-acting tidal wave. X chromosome to autosome translocations and incorporation of transgenes containing *XIST* into autosomal DNA show that inactivation can often spread from the X inactivation center (XIC) beyond X chromosomal DNA, into random continguous DNA. But examples of the forceful spread of X inactivation probably do not reflect how X inactivation actually evolved. After all, cells with X to autosome translocations or random *XIST* incorporation tend to

die. For example, in mice with balanced X to chromosome 16 Searle's translocations, only the cells that inactivate the intact X chromosome survive; apparently none of the cells survive that inactivate the chromosome comprised of the XIC-carrying portion of the X attached to a chromosome 16 fragment (Takagi, 1980). When a transgene containing the XIC was permitted to incorporate randomly into DNA, excessive cell death was observed in ES cells that were permitted to differentiate (Lee and Jaenisch, 1996). So, though X to autosome translocations and XIC incorporation into autosomes suggest that X inactivation is a sweeping process, acquiring inactivation in great sweeps into unknown territories also appears to be lethal and may be far from the way X inactivation evolved.

The data indicate that in recent mammalian evolution, X inactivation has evolved gene by gene as Y genes have decayed. Perhaps this is the way X inactivation has always evolved: gene by gene, ever since the Y was defined and genes on it began to decay. If this is so, the mechanism of X inactivation may operate at a finer level than often supposed, acting a patch at a time. Perhaps X inactivation has come to look like one sweeping process because the patchwork acquisition of inactivation has become so complete as to seem seamless, except on painstaking inspection.

References

Bull, J. J. (1983). *Evolution of Sex Determining Mechanisms*, 1st Edition, D. B. Wake and M. W. Slatkin, eds. (Menlo Park, CA: Benjamin/Cummings).

Carrel, L., Hunt, P. A., and Willard, H. F. (1996). Tissue and lineage-specific variation in inactive X chromosome expression of the murine Smcx gene. *Hum Mol Genet 5*, 1361-1366.

Clemson, C. M., McNeil, J. A., Willard, H. F., and Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J Cell Biol 132*, 259-275.

Ferguson-Smith, M. A. (1965). Karyotype-phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations. *J Med Genet 2*, 142-155.

Graves, J. A. (1987). The evolution of mammalian sex chromosomes and dosage compensation — clues from marsupials and monotremes. *Trends Genet 3*, 252-256.

Graves, J. A. (1995). The evolution of mammalian sex chromosomes and the origin of sex determining genes. *Philos Trans R Soc Lond B Biol Sci 350*, 305-312.

Graves, J. A., and Watson, J. M. (1991). Mammalian sex chromosomes: evolution of organization and function. *Chromosoma 101*, 63-68.

Grumbach, M. M., Morishima, A., and Taylor, J. H. (1963). Human sex chromosome abnormalities in relation to DNA replication and heterochromatinization. *Proc Natl Acad Sci USA 49*, 581-589.

Jeppesen, P., and Turner, B. M. (1993). The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell 74*, 281-289.

Lee, J.T., Strauss, W.M., Dausman, J.A. and Jaenisch, R. (1996). A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell 86*, 83-94.

Meroni, G., Franco, B., Archidiacono, N., Messali, S., Andolfi, G., Rocchi, M., and Ballabio, A. (1996). Characterization of a cluster of sulfatase genes on Xp22.3 suggests

gene duplications in an ancestral pseudoautosomal region. *Hum Molec Genet 5*, 423-431.

Migeon, B. R., Shapiro, L. J., Norum, R. A., Mohandas, T., Axelman, J., Dabora, R. L. (1982). Differential expression of steroid sulfatase locus on active and inactive human X chromosome. *Nature 299*, 838-840.

Odorisio, T., Mahadevaiah, S. K., McCarrey, J. R., and Burgoyne, P. S. (1996). Transcriptional analysis of the candidate spermatogenesis gene Ube1y and of the closely related Ube1x shows that they are coexpressed in spermatogonia and spermatids but are repressed in pachytene spermatocytes. *Develop Biol 180*, 336-343.

Sheardown, S., Norris, D., Fisher, A., and Brockdoff, N. (1996). The mouse Smcx gene exhibits developmental and tissue specific variation in degree of escape from X inactivation. *Hum Mol Genet 5*, 1355-1360.

Takagi, N. (1980). Primary and secondary nonrandom X chromosome inactivation in early female mouse embryos carrying Searle's translocation T(X;16)16H. *Chromosoma 81*, 439-459.

Walker, C. L., Cargile, C. B., Floy, K. M., Delannoy, M., and Migeon, B. R. (1991). The Barr body is a looped X chromosome formed by telomere association. *Proc Natl Acad Sci USA 88*, 6191-6195.

Appendix:

Documenting Additional Cases

DBX/DBY

An X-Y homologous pair of RNA helicase genes was recently identified (Lahn and Page, unpublished data) which encode "DEAD box" type proteins. These X-Y homologous genes, like others previously identified, are similar in their coding region (88% nucleotide identity, 91% amino acid identity) but diverged in flanking, untranslated sequences, except notably over a long stretch following the stop codon in these genes' long 3' untranslated regions. A close homolog of *DBX* is known in mouse; it has not been shown to map to the X chromosome but very likely does given the almost universal conservation of the set of X-linked genes in different placental mammals. Bruce Lahn showed that the 5' region of human *DBX* is not methylated in females, implying that the X-Y homologous gene escapes X inactivation.

Sequence near the start codon of *DBX* is well conserved between humans and mice (figure A.1) and has the quality of a CpG island. Indeed, the sequence is so well conserved that PCR primers designed to be complementary to both human and mouse also amplified products from marsupial DNA (see opossum in figure A.2). The 66 bp PCR product includes a GCGC site, the recognition site for the methylation sensitive restriction enzyme, HhaI.

Genomic DNA was digested either with HindIII, which does not cut between the PCR primers, or with HhaI before PCR amplification. Amplification is reduced after HhaI digestion if the template is unmethylated. HhaI, unlike HpaII with MspI, does not have a methylation insensitive isoschizomer. But male DNA can serve as a control to show what unmethylated DNA should look like.

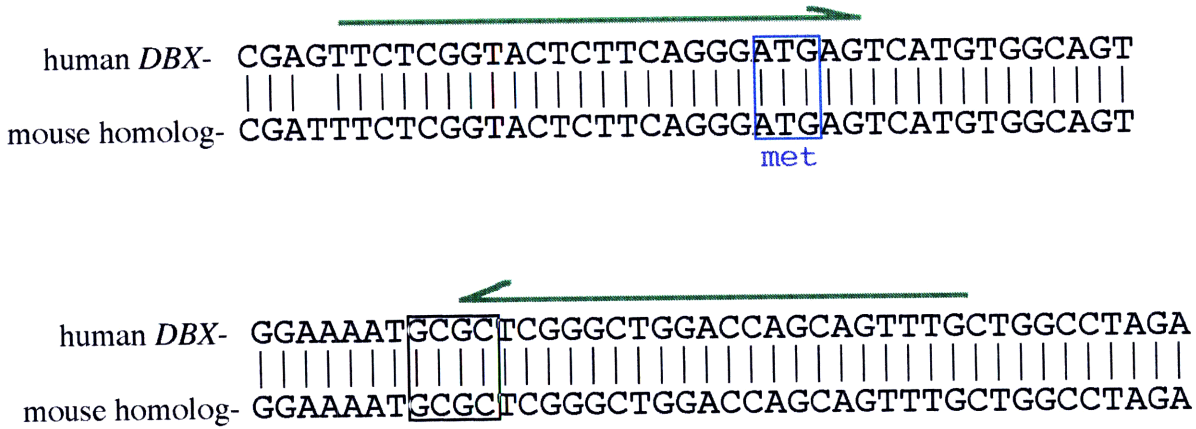


Figure A.1 Alignment of human *DBX* and its close homolog, mouse RNA helicase, in the 5' CpG island region. The tested HhaI (GCGC) site is boxed. PCR primers used to assay methylation and the start codon are also indicated. The CpG to GpC ratio in this region is about 0.9, and therefore CpG island-like.

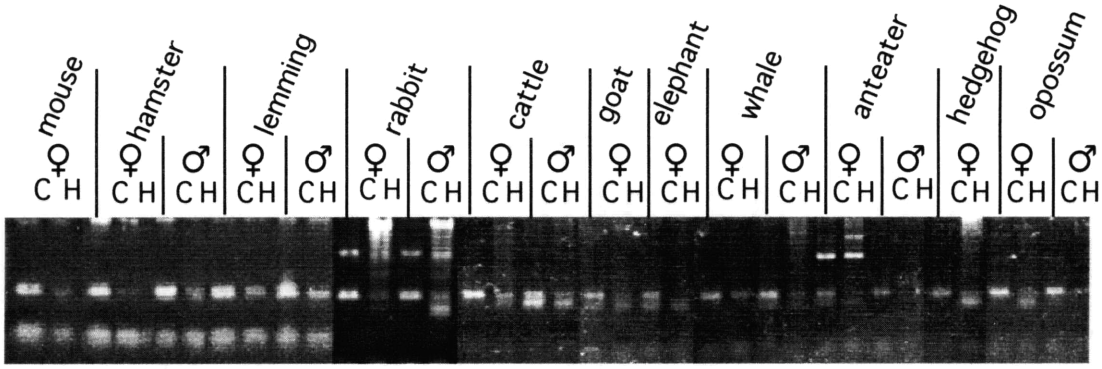


Figure A.2 *DBX* methylation results. Female and male genomic DNAs digested with HhaI (H) or HindIII (C, "control") were used as alternate templates in PCR assay diagrammed in figure A.1. PCR products were visualized by ethidium bromide/UV staining after agarose gel electrophoresis.

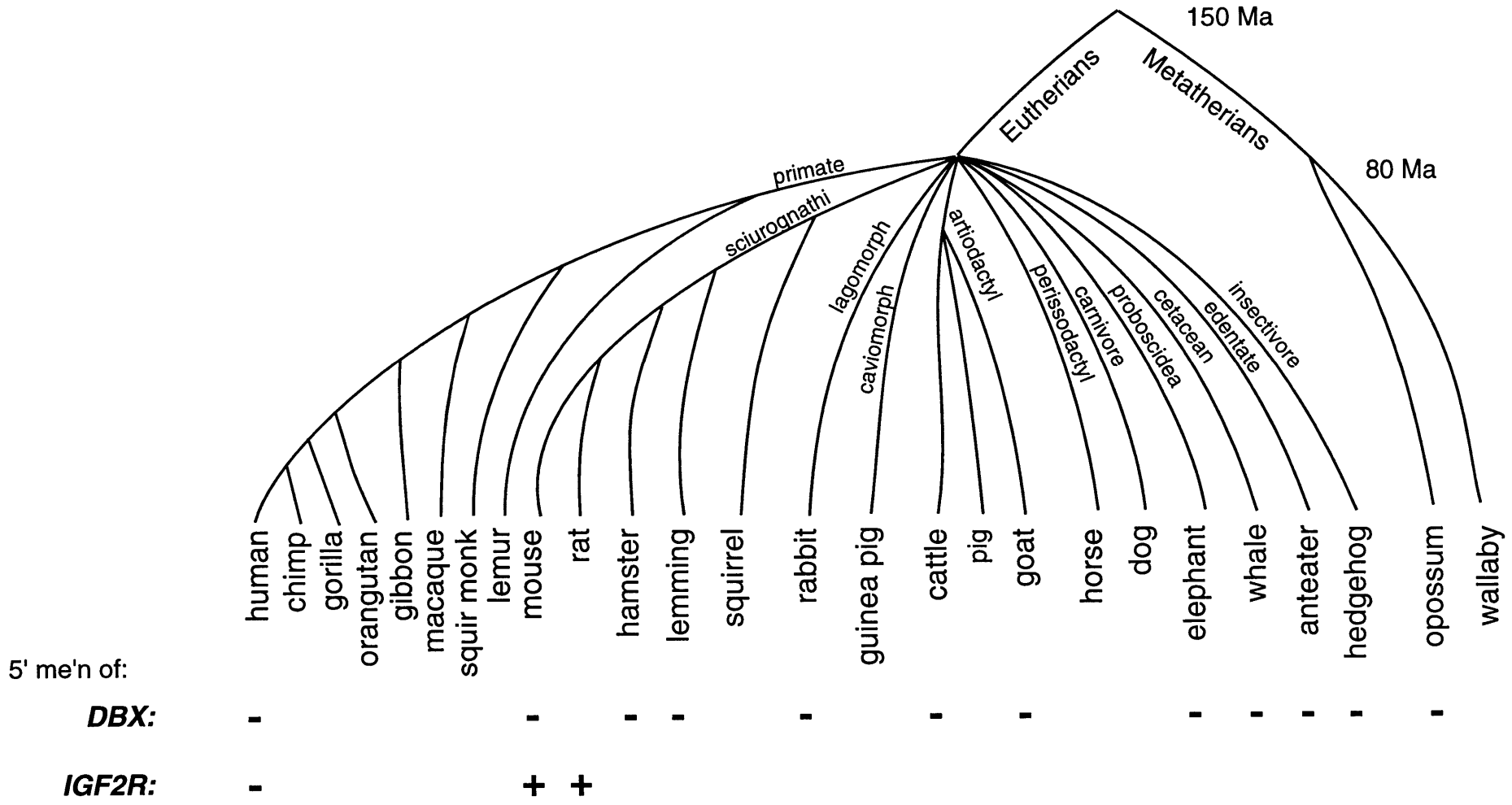


Figure A.3 Results of methylation studies for *DBX*, an X-Y homologous gene, and for *IGF2R*, a gene imprinted in mice but not in humans, are shown in phylogenetic context.

DBX appears unmethylated in females of 11 eutherian species representing 9 orders. In no species does it appear methylated. In all testable eutherians, *DBX* appears to escape X inactivation (figures A.2, A.3). Every X-Y homologous gene investigated has presented a unique pattern of X inactivation when examined in a wide context. The theme emerges from studying exceptional genes that X inactivation can be acquired in a punctile way.

Methylation assay. Conditions were as described for *ZFX* in Chapter 2 with the following modifications: genomic DNA was digested with HindIII and HhaI; PCR was performed with the annealing temperature of 60°C. The primers used in the PCR assay were TTCTCGGTACTCTTCAGGGATGAG and AAACTGCTGGTCCAGCCCGAGC and amplified a 67 bp product that included one HhaI site.

IGF2R

Mammalian genomic imprinting—the phenomenon where a gene's expression depends on the parent from which it is inherited—represents, like X inactivation, a case of allelic inactivation. Several genes, like *IGF2* and *H19*, are imprinted in mice and in humans. *IGF2R*, which encodes the insulin-like growth factor type 2 receptor, shows differential imprinting between humans and mice. It is not imprinted in humans; both alleles of *IGF2R* are expressed. In mice, the gene is imprinted; the paternal allele is repressed, and the maternally inherited allele is expressed (Kalscheuer et al., 1993).

Two methylated regions have been found in mouse *Igf2r* (Stoger et al., 1993). A CpG island exists within the gene's second intron, and its methylation acts as the imprinting signal; only the maternal allele, fated to be expressed, is methylated there in mouse. The 5' region also contains a CpG island, and in mouse the silent paternal allele is methylated there. Methylation of the 5' region of *IGF2R* (figure A.4) appears to serve a purpose analogous to methylation of genes on the inactive X, a way to cement allelic inactivation.

The mechanism of *IGF2R* imprinting remains to be thoroughly promulgated. The CpG island within the mouse gene's second intron appears to contain the imprinting signal: yeast artificial chromosome transgenes reproduced the imprinted methylation and expression patterns of *Igf2r* as long as the second intron was undisturbed (Wutz, et al., 1997). Yet a CpG island exists in the human *IGF2R*'s second intron as well, and that island has been reported to be preferentially methylated on the maternal allele too (Riesewijk, et al., 1996) even though the expression of one allele is not favored in humans. Whatever the signal that leads to imprinting, and that original signal apparently involves the CpG island in intron 2, methylation of the 5' CpG island appears to be a reliable marker for allelic inactivation (Riesewijk, et al., 1996).

Human *IGF2R*

gcccgcgccgcttcctgtcccgcgcgctgcgctcacgtgacccgggcctgggagg

agccggggcgggcgggggtcacctgaacaagggagtcacgtgagcgggggggcggg

ggtgggggggcggtgccgggcggctgtcacgtgacgcggttccggggccgccgct

gccgctgtcgctgtcgccgagcccagtcgagccgcgctcacctcgggctcccgct

ccgtctccacctccgcctttgccctggcggcgcgaccccgtcccgggcgcggccc

ccagcagtcgcgcgccgttagcctcgcgcccgccgcgcagctccgggcccggcgc

gATGggggccgccgccggccggagcccccacctggggcccgcgcccgcccgccgc

met

ccgcagcgctctctgctcctgctgcagctgctgctgctcgtcgctgccccgggggt

ccacgcaggcccaggccgccccgttccccgagctgtgcaggtgggtggcc

Mouse *Igf2r*

cctcgagtaggtacctggcgctcgtgccccgggcccgcaacacttcctgtcccgcgc

gcgtgcgatgctcatgtgaccccgggactgggcggagagcacctgaacgaggacgt

cacgtgagcaggaggcggggcgggggcgggccgactcaggtcacgtgacgctccg

gggacggccacggagcgcctcctcgtcgcactccccctggctccagttctctct

cctctttctccctccagctcccgttgcagcttcgactccgctgtggtggcgcgac

cgtgtcccaggcgcggctccaaacggccagccgccgtgagccccacgccacacgc

gATGcgggccgttcagctgggaccggtgccctccgggccgcgcgtcgcgctcctg

met

ccgccgctcctgctgctgctcctgctggcggccgcgggctccgcgcaggcccagg

ccgtcgacttggacgccctgtgcaggtgggtgtcttgcgt

Figure A.4 Design of methylation assays for human and mouse *IGF2R*. Tested HpaII/MspI (CCGG) sites and PCR primers used are indicated. The start codon is also shown in each case.

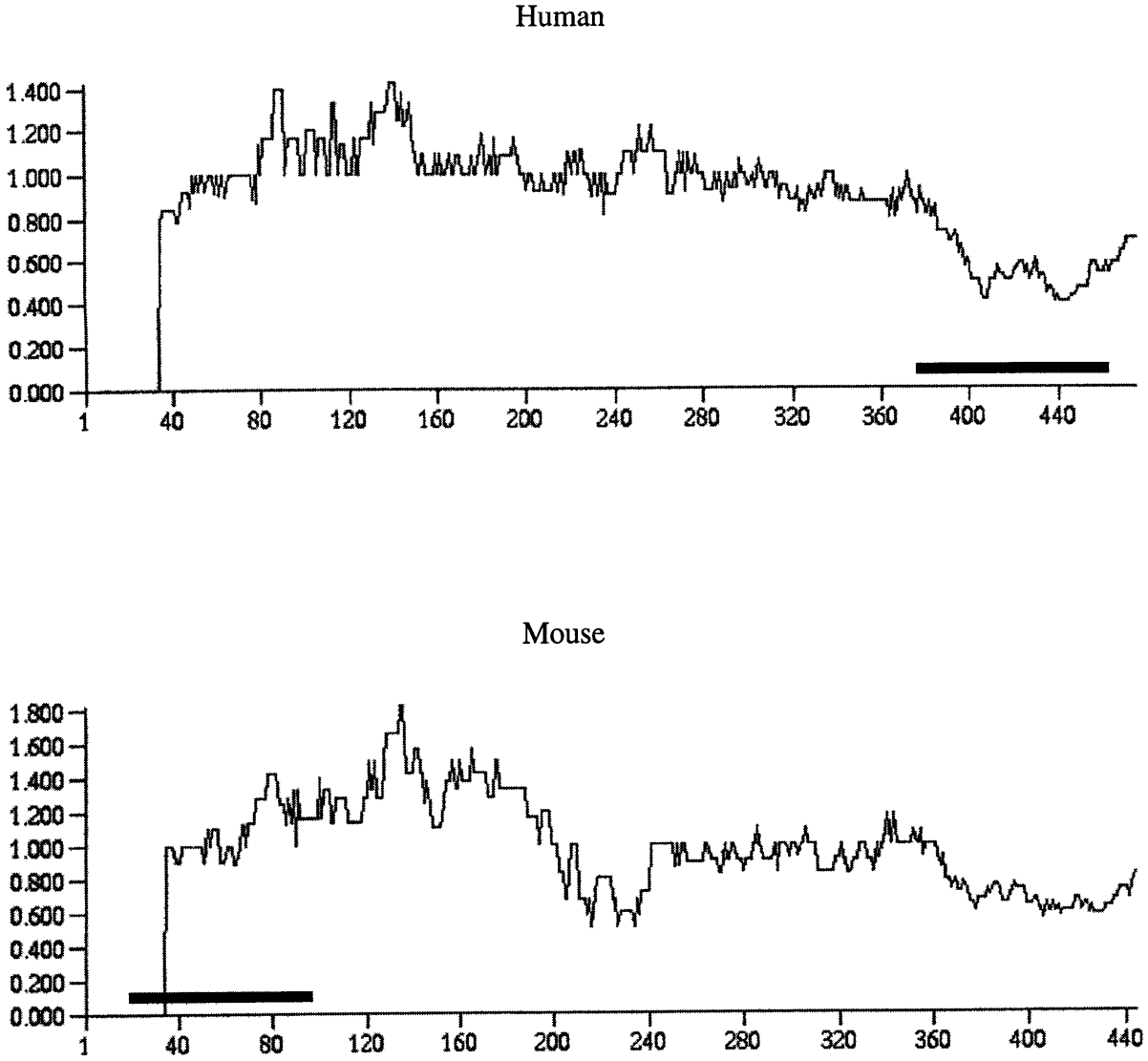


Figure A.5 CpG to GpC dinucleotide ratios in the 5' regions of human and mouse *IGF2R*, where the assays shown in figure A.4 were designed. In mammalian genomic DNA the CpG to GpC ratio tends to be near 0.2 but in CpG islands approaches 1. Ratios were calculated over 70 bp windows. Areas where methylation was tested are marked with bars over the *x* axes.

Note that discrepancies are expected at the 5' and 3' ends of these kinds of graphs because dinucleotide counting occurs over windows.

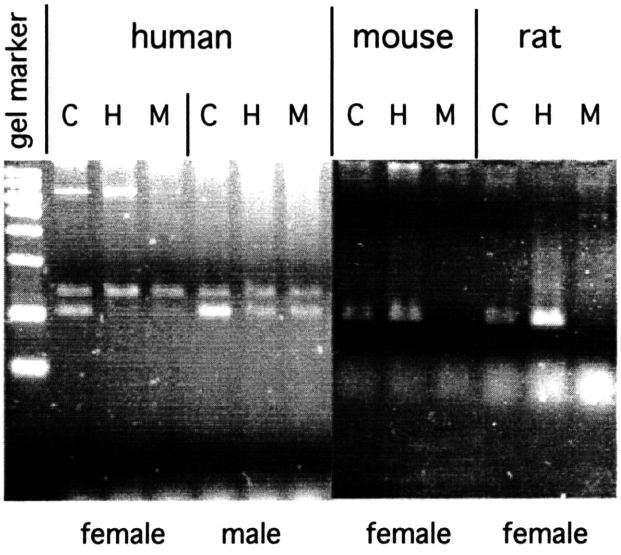


Figure A.6 *IGF2R* methylation results. Genomic DNAs digested with HpaII (H) or MspI (M) or HindIII (C, "control") were used as alternate templates in PCR assay diagrammed in figure A.4. PCR products were visualized by ethidium bromide/UV staining after agarose gel electrophoresis.

The 5' region of IGF2R is poorly conserved and exceptionally CG rich (the latter property can foil the design of workable PCR reactions) (figures A.4, A.5). A PCR reaction designed to flank CCGG restriction enzyme sites confirmed that the 5' region *Igf2r* is methylated in mice and rat. (One allele, the paternal, is expected to be methylated in a mouse, no matter what its sex.) A different assay in a similar region of human *IGF2R* showed that the promoter area of the gene is not methylated in humans, as expected given the absence of *IGF2R* imprinting in humans (figure A.6).

These results show that 5' methylation correlates with the presence of an inactivated allele in the case of genomic imprinting, just as with X inactivation.

Methylation assay. Again, conditions were as described in Chapter 2. The human assay used an annealing temperature of 62°C and the primers GCCCGCAGCGCTCTCT and GCCACCCACCTGCACAG to amplify a 105 bp product including one HpaII/MspI site. The mouse/rat assay used an annealing temperature of 60°C and the primers GTAGGTACCTGGCGCTCG and GTGACGTCCTCGTTCAGGTG to amplify a 106 bp product including two HpaII/MspI sites.

References

Kalscheuer, V. M., Mariman, E. C., Schepens, M. T., Rehder, H., and Ropers, H. H. (1993). The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat Genet 5*, 74-78.

Riesewijk, A. M., Schepens, M. T., Welch, T. R., van den Berg-Loonen, E. M., Mariman, E. M, et al. (1996). Maternal-specific mehtylation of the human IGF2R gene is not accompanied by allele-specific transcription. *Genomics 31*, 158-166.

Stoger, R., Kubicka, P., Liu, C. G., Kafri, T., Razin, A., Cedar, H., and Barlow, D. P. (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell 73*, 61-71.

Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F., Barlow, D. P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature 389*, 745-747.